

94473

From: Rao, Manjunath N.  
Sent: Tuesday, May 20, 2003 11:54 AM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search request for 09/869,155

From: Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10D 01  
Phone: 306-5681

Point of Contact:  
Toby Port  
Technical Info. Specialist  
CM1 6A04  
703-308-3534

Date: 5-20-03

Please search the following as soon as possible for application with serial number **09/869,155**

1. SEQ ID NO:13, 14, 15, 16, 17, 18, 19 against all commercial protein databases including issued patents database and pending application database and provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.  
Biotechnology Patent Examiner  
Art Unit 1652, Room 10A11  
Mail Box in 10D01  
Crystal Mall 1, USPTO.

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 5/20  
Date Completed: 5/20  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:49:49 ; Search time 8.40659 Seconds  
(without alignments)  
200.549 Million cell updates/sec

Title: US-09-869-155-14

Perfect score: 96  
Sequence: 1 KRLGFSRLPHFTGCGL 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published\_Applications\_AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCOT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCOTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44.5	46.4	218	9	US-09-738-626-4068
2	43	44.8	163	9	US-10-211-962-76
3	42	43.8	163	9	US-10-211-962-71
4	41	42.7	163	9	US-10-211-962-72
5	41	42.7	163	9	US-10-211-962-75
6	41	42.7	729	9	US-10-051-902-29
7	41	42.7	729	9	US-10-051-909-29
8	40	41.7	163	9	US-10-211-962-79
9	40	41.7	324	10	US-09-816-095-4
10	40	41.7	341	9	US-10-043-487-290
11	40	41.7	372	9	US-09-860-670-106
12	39	40.6	59	9	US-09-796-692-1038
13	39	40.6	59	9	US-10-040-862-1038
14	39	40.6	64	9	US-10-150-111-155
15	39	40.6	572	9	US-10-272-419-8
16	38	39.6	25	9	US-10-023-282-454
17	38	39.6	85	9	US-10-114-893-194
18	38	39.6	123	10	US-09-867-550-1202
19	38	39.6	197	10	US-09-764-853-774

20	38	39.6	363	9	US-09-738-626-5456	Sequence 5456, App
21	38	39.6	499	10	US-09-976-165-40	Sequence 40, App1
22	38	39.6	620	10	US-09-925-301-1193	Sequence 1193, Ap
23	38	39.6	737	9	US-10-051-902-8	Sequence 8, App11
24	38	39.6	737	9	US-10-051-909-8	Sequence 8, App11
25	38	39.6	782	9	US-10-097-340-312	Sequence 312, App
26	38	39.6	855	9	US-10-099-700A-2	Sequence 2, App11
27	38	39.6	1189	9	US-09-738-626-4140	Sequence 4140, Ap
28	38	39.6	5215	9	US-09-860-846-2	Sequence 2, App11
29	38	39.6	5215	9	US-09-888-384B-2	Sequence 2, App11
30	38	39.6	5215	9	US-09-836-821-2	Sequence 2, App11
31	38	39.6	5215	10	US-09-861-289-2	Sequence 2, App11
32	37.5	39.1	124	9	US-09-764-891-4424	Sequence 4424, Ap
33	37.5	39.1	402	10	US-09-815-242-5748	Sequence 5748, A
34	37.5	39.1	408	10	US-09-815-242-12296	Sequence 12296, A
35	37.5	39.1	669	9	US-10-024-632-9	Sequence 9, App11
36	37	38.5	67	9	US-09-796-692-752	Sequence 752, App
37	37	38.5	67	9	US-10-040-862-752	Sequence 752, App
38	37	38.5	80	9	US-09-796-692-889	Sequence 889, App
39	37	38.5	80	9	US-10-040-862-889	Sequence 889, App
40	37	38.5	120	9	US-09-796-692-807	Sequence 807, App
41	37	38.5	120	9	US-10-040-862-807	Sequence 807, App
42	37	38.5	295	10	US-09-815-242-11239	Sequence 11239, A
43	37	38.5	567	10	US-09-815-242-13511	Sequence 13511, A
44	37	38.5	740	9	US-10-051-909-37	Sequence 37, App1
45	37	38.5	800	9	US-10-051-909-32	Sequence 32, App1

#### ALIGNMENTS

RESULT 1  
US-09-738-626-4068  
; Sequence 4068, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAMA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4068  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4068

Query Match 46.4%; Score 44.5; DB 9; Length 218;  
Best Local Similarity 62.5%; Pred. No. 17;  
Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;  
OY 2 KRLGFSRLPHFTGCGL 17  
DB 76 RCGSIRAPHFTG-GGI 90

```
RESULT 2
US-10-211-962-76
; Sequence 76, Application US/10211962
; Publication No. US20030082640A1
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/10/211,962
; CURRENT FILING DATE: 2002-08-01
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 76
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-211-962-76

Query Match          44.8%; Score 43; DB 9; Length 163;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KRIGFSRLPHFT 12
Db 84 KMGFGRLHFTS 95

RESULT 3
US-10-211-962-71
; Sequence 71, Application US/10211962
; Publication No. US20030082640A1
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/10/211,962
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US/09/562,737
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-211-962-71

Query Match          43.8%; Score 42; DB 9; Length 163;
Best Local Similarity 61.5%; Pred. No. 31;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KRIGFSRLPHFTG 13
Db 84 KMGFGRLHFTSG 96

RESULT 4
US-10-211-962-72
; Sequence 72, Application US/10211962
; Publication No. US20030082640A1
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
```

```
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/10/211,962
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US/09/562,737
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-211-962-72

Query Match          42.7%; Score 41; DB 9; Length 163;
Best Local Similarity 61.5%; Pred. No. 44;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KRIGFSRLPHFTG 13
Db 84 KMGFGRLHFTSG 96

RESULT 5
US-10-211-962-75
; Sequence 75, Application US/10211962
; Publication No. US20030082640A1
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/10/211,962
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US/09/562,737
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 75
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-211-962-75

Query Match          42.7%; Score 41; DB 9; Length 163;
Best Local Similarity 61.5%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 KRIGFSRLPHFTG 13
Db 84 KMGFGRLHFTSG 96

RESULT 6
US-10-051-902-29
; Sequence 29, Application US/10051902
; Patent No. US20020178468A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Hiltz, Bill
; APPLICANT: Kinney, Tony
; APPLICANT: Tingey, Scott
; TITLE OF INVENTION: Plant Sugar Transport Proteins
; FILE REFERENCE: BB-1163
; CURRENT APPLICATION NUMBER: US/10/051,902
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/291,922
```

```

; PRIOR FILING DATE: EARLIER FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 29
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-051-902-29

Query Match
Best Local Similarity 42.7%; Score 41; DB 9; Length 729;
Best Local Similarity 70.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 LPFTGCGGL 17
DB 136 LPFTGCGGM 145

RESULT 7
US-10-051-909-29
; Sequence 29, Application US/10051909
; Publication No. US20020199217A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Heltjars, Tim
; APPLICANT: Hiltz, Bill
; APPLICANT: Kinney, Tony
; APPLICANT: Tinney, Scott
; TITLE OF INVENTION: Plant Sugar Transport Proteins
; FILE REFERENCE: Bb1163 US-CIP
; CURRENT APPLICATION NUMBER: US/10/051,909
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 60/083,044
; PRIOR FILING DATE: April 24, 1998
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 29
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-051-909-29

Query Match
Best Local Similarity 42.7%; Score 41; DB 9; Length 729;
Best Local Similarity 70.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 LPFTGCGGL 17
DB 136 LPFTGCGGM 145

RESULT 8
US-10-211-962-79
; Sequence 79, Application US/10211962
; Publication No. US20030082640A1
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/10/211,962
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US/09/562,737
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 79
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
```

```

US-10-211-962-79

Query Match
Best Local Similarity 41.7%; Score 40; DB 9; Length 163;
Best Local Similarity 61.5%; Pred. No. 64;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KRGFSLPHTG 13
DB 84 KMGFLRLHFSG 96

RESULT 9
US-09-816-095-4
; Sequence 4, Application US/09816095
; Patent No. US20020137164A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CL001147
; CURRENT APPLICATION NUMBER: US/09/816,095
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-816-095-4

Query Match
Best Local Similarity 41.7%; Score 40; DB 10; Length 324;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 PHTGCGG 16
DB 42 PHTVCGG 49

RESULT 10
US-10-043-487-290
; Sequence 290, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; APPLICANT: Pierre, DEGRAIN
; TITLE OF INVENTION: Protein-protein interactions between Shigella flexneri polypep
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043,487
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 290
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Shigella flexneri
US-10-043-487-290

Query Match
Best Local Similarity 41.7%; Score 40; DB 9; Length 341;
Best Local Similarity 57.1%; Pred. No. 1.3e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 GFSRLPHTGCGGL 17
DB 291 GTSRLPHASSTGMD 304

RESULT 11
US-09-860-670-106
```

```
; Sequence 106, Application US/09860670
; Patent No. US20020165137A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA127PI
; CURRENT APPLICATION NUMBER: US/09/860,670
; PRIOR FILING DATE: 2001-05-21
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 106
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-860-670-106

Query Match          41.7%; Score 40; DB 9; Length 372;
Best Local Similarity 47.8%; Pred. No. 1.4e+02;
Matches 11; Conservative 2; Mismatches 4; Indels 6; Gaps 2;

QY 1 KRIGFSR---LPHF---TGGCGL 17
   11:11:1 1 1 11 11
Db 56 KRIGFMSYSLASFIFAAGCAGL 78

RESULT 12
US-09-796-692-1038
; Sequence 1038, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Aigate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1038
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-796-692-1038

Query Match          40.6%; Score 39; DB 9; Length 59;
Best Local Similarity 58.3%; Pred. No. 34;
```

```
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 LGFSRLPHFTGC 14
   11:11111 1
Db 18 LGSVAVLPHFPSC 29

RESULT 13
US-10-040-862-1038
; Sequence 1038, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Aigate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1038
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-040-862-1038

Query Match          40.6%; Score 39; DB 9; Length 59;
Best Local Similarity 58.3%; Pred. No. 34;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 LGFSRLPHFTGC 14
   11:11111 1
Db 18 LGSVAVLPHFPSC 29

RESULT 14
US-10-150-111-155
; Sequence 155, Application US/10150111
; Publication No. US20030078386A1
; GENERAL INFORMATION:
; APPLICANT: Rubin et al.
; TITLE OF INVENTION: Secreted Protein HPEAD48
; FILE REFERENCE: P2018P1D1
```

;; CURRENT APPLICATION NUMBER: US/10/150,111  
;; CURRENT FILING DATE: 2002-05-20  
;; PRIOR APPLICATION NUMBER: 09/288,143  
;; PRIOR FILING DATE: 1999-04-08  
;; PRIOR APPLICATION NUMBER: PCT/US98/21142  
;; PRIOR FILING DATE: 1998-10-08  
;; PRIOR APPLICATION NUMBER: 60/061,463  
;; PRIOR FILING DATE: 1997-10-09  
;; PRIOR APPLICATION NUMBER: 60/061,529  
;; PRIOR FILING DATE: 1997-10-09  
;; PRIOR APPLICATION NUMBER: 60/071,498  
;; PRIOR FILING DATE: 1997-10-09  
;; PRIOR APPLICATION NUMBER: 60/061,527  
;; PRIOR FILING DATE: 1997-10-09  
;; PRIOR APPLICATION NUMBER: 60/061,536  
;; PRIOR FILING DATE: 1997-10-09  
;; PRIOR APPLICATION NUMBER: 60/061,532  
;; PRIOR FILING DATE: 1997-10-09  
;; NUMBER OF SEQ ID NOS: 219  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO: 155  
;; LENGTH: 64  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-150-111-155

Query Match 40.6%; Score 39; DB 9; Length 64;  
Best Local Similarity 47.6%; Pred. No. 37;  
Matches 10; Conservative 2; Mismatches 5; Indels 4; Gaps 1;

OY 1 KRIGSRPHFT---GCGGL 17  
I :||| ||:|  
Db 16 KWIGSFPHWTWIDELIGL 36

RESULT 15  
US-10-272-419-8  
;; Sequence 8, Application US/10272419  
;; Publication No. US20030087403A1  
;; GENERAL INFORMATION:  
;; APPLICANT: CHEN, QIONG  
;; APPLICANT: THOMAS, STUART  
;; APPLICANT: NAGARAJAN, VASANTHA  
;; TITLE OF INVENTION: BIOLOGICAL METHOD FOR THE PRODUCTION OF ADIPIC ACID AND  
;; FILE REFERENCE: C11341-A  
;; CURRENT APPLICATION NUMBER: US/10/272,419  
;; CURRENT FILING DATE: 2002-10-16  
;; PRIOR APPLICATION NUMBER: 09/252,553  
;; PRIOR FILING DATE: 1999-02-19  
;; NUMBER OF SEQ ID NOS: 32  
;; SOFTWARE: Microsoft Office 97  
;; SEQ ID NO: 8  
;; LENGTH: 572  
;; TYPE: PRT  
;; ORGANISM: Acinetobacter sp.  
US-10-272-419-8

Query Match 40.6%; Score 39; DB 9; Length 572;  
Best Local Similarity 61.5%; Pred. No. 31e+02;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 4 GFSRPHFTGCG 16  
I :||| ||:|  
Db 316 GFNSLPKFTFPBG 328

Search completed: May 20, 2003, 18:00:00  
Job time : 8.40659 secs

**THIS PAGE BLANK (USPTO)**

PT Mutant xylanase protein identified using xylanase inhibitor useful for  
PT preparing non-sticky dough for bakery products -

XX Claim 24: Page 112; 112pp; English.  
PS  
XX The present sequence is derived from an endo-beta-1,4-xyylanase  
CC inhibitor. The protein is obtained from wheat flour. The specification  
CC also describes a mutant xyylanase protein. The xyylanase is useful for  
CC preparing a foodstuff, preferably a bakery product or a substance  
CC (e.g. a dough) for making the bakery product. Wild type xyylanase or  
CC mutant xyylanase is useful for preparing a dough that is less sticky  
CC than a dough comprising a fungal xyylanase. The xyylanase inhibitor is  
CC useful for screening high degree resistance xyylanses for dough  
CC preparation. The xyylanase is also useful for preparing a non-sticky  
CC dough. A combination of xyylanase and the inhibitor is useful for  
CC calibrating and/or determining the quantity of inhibitor in a wheat  
CC flour sample.  
XX  
SQ Sequence 35 AA:  
  
Query Match 100.0%; Score 185; DB 21; Length 35;  
Best Local Similarity 100.0%; Pred. No. 7.3e-21;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
2Y 1 GAPVARAVEAVAPFGVCYDTKTLGNLGGYAVPNV 35  
Db 1 GAPVARAVEAVAPFGVCYDTKTLGNLGGYAVPNV 35  
|||||  
  
RESULT 2  
AAE14687  
ID AAE14687 standard: peptide; 35 AA.  
XX  
XX AAE14687;  
AC  
XX  
XX 21-AUG-2002 (first entry)  
DT  
XX  
XX Wheat flour xyylanase inhibitor A chain N-terminal fragment.  
DE  
XX Refrigerated dough; syrruping; arabinoxylan; bakery product; bread;  
XX pizza base; cake; biscuit; wheat; flour; xyylanase inhibitor.  
KM  
XX Triflicum aestivum.  
OS  
XX WO200152657-A1.  
PI  
XX 26-JUL-2001.  
XX  
XX 17-JAN-2001; 2001WO-1B00168.  
XX  
XX 18-JAN-2000; 2000GB-0001136.  
XX  
XX (DANI-) DANISCO AS.  
PA  
XX  
XX Poulsen CH, Sorensen JF;  
PI  
XX WPI: 2001-457446/49.  
XX  
XX Production of refrigerated dough with reduced syrruping, useful in  
PT production of bakery products such as bread, comprises admixing cereal  
PT flour, water and protein that prevents enzymatic degradation of  
PT arabinoxylan in the cereal flour -  
XX  
XX Disclosure: Page 23; 26pp; English.  
PS  
XX The invention relates to a process for producing refrigerated dough  
CC with reduced 'syruping' (precipitation of liquid on the dough surface  
CC because of a reduction in water holding capacity caused by the breakdown  
CC of arabinoxylan over time). The process comprises admixing cereal flour  
CC and water with a protein that reduces/prevents enzymatic degradation of  
CC arabinoxylan in the cereal flour. The preferred protein is a xyylanase  
CC inhibitor. The method is useful to produce refrigerated dough in which  
CC syruping is reduced or eliminated. Refrigerated dough is typically  
CC stored for long periods to enable fresh baked products (e.g. bread,  
CC pizza bases, cakes, biscuits) to be produced as required. It may also be

CC possible to produce dough or baked products tailored to specific  
CC requirements using the method by the use of specific proteins/protein  
CC combinations. The present sequence is wheat flour  
CC endo-beta-1,4-xyylanase inhibitor A chain N-terminal fragment.  
XX  
SQ Sequence 35 AA:  
  
Query Match 100.0%; Score 185; DB 22; Length 35;  
Best Local Similarity 100.0%; Pred. No. 7.3e-21;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GAPVARAVEAVAPFGVCYDTKTLGNLGGYAVPNV 35  
Db 1 GAPVARAVEAVAPFGVCYDTKTLGNLGGYAVPNV 35  
|||||  
  
RESULT 3  
AAU07392  
ID AAU07392 standard: protein; 35 AA.  
XX  
XX AAU07392;  
AC  
XX  
XX 18-DEC-2001 (first entry)  
DT  
XX  
XX Bacillus subtilis xyylanase inhibitor #1.  
DE  
XX xyylanase; plant cell wall; baking; cereal; starch production; wood;  
XX wood pulp bleaching; hemicellulose; arabinoxylan; foodstuff; inhibitor.  
KM  
XX Bacillus subtilis.  
OS  
XX WO200166711-A1.  
PI  
XX  
XX 13-SEP-2001.  
XX  
XX 08-MAR-2001; 2001WO-1B00426.  
XX  
XX 08-MAR-2000; 2000GB-0005585.  
XX  
XX 27-JUN-2000; 2000GB-0015751.  
XX  
XX (DANI-) DANISCO AS.  
PA  
XX  
XX Sibbesen O, Sorensen JF;  
PI  
XX WPI: 2001-596834/67.  
XX  
XX Novel variant xyylanase polypeptide or its fragment useful for degrading  
PT or modifying plant cell wall, comprises amino acid modifications such  
PT that the polypeptide has altered sensitivity to xyylanase inhibitor -  
XX  
XX Disclosure: Page 63; 70pp; English.  
PS  
XX The invention relates to a variant xyylanase polypeptide (I) or its  
CC fragment having xyylanase activity, comprising one or more amino acid  
CC modifications such that (I) or its fragment has an altered sensitivity to  
CC a xyylanase inhibitor as compared with the parent xyylanase enzyme. (I) or  
CC its coding sequence (II) is useful for degrading or modifying plant cell  
CC wall or for processing a plant material by contacting the plant cell wall  
CC or plant material with (I) or (II). (I) is useful for modifying plant  
CC materials, and in baking, processing cereals, starch production,  
CC processing wood and enhancing the bleaching of wood pulp. (I)  
CC is useful for altering the viscosity derived from the presence of  
CC hemicellulose or arabinoxylan in a solution or system comprising plant  
CC cell wall material. (I) is useful for preparing a foodstuff such as  
CC bread, pretzels, tortillas, cakes, cookies, biscuits or crackers.  
CC The present sequence represents the amino acid sequence of Bacillus  
CC subtilis xyylanase inhibitor #1 as described in the method of the  
CC invention.  
XX  
SQ Sequence 35 AA:  
  
Query Match 100.0%; Score 185; DB 22; Length 35;  
Best Local Similarity 100.0%; Pred. No. 7.3e-21;



Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAPVARAEAVAPFGVCYDRTKLTGNNLGGYAVPNV 35  
 |||||||||||||||||||||||||||||||||||

DB 1 GAPVARAEAVAPFGVCYDRTKLTGNNLGGYAVPNV 35

RESULT 4  
 AAU75817  
 ID AAU75817 standard; Protein: 196 AA.

XX AC AAU75817;  
 XX DT 23-APR-2002 (first entry)  
 XX DE Wheat L endoxylanase inhibitor, TDXI I, partial sequence TDXI-I.01.  
 XX KW Wheat; TDXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX OS Triticum durum cultivar Mexicali.  
 XX PN MO200198474-A1.  
 XX PD 27-DEC-2001.  
 XX PF 21-JUN-2001; 2001WO-BE00106.  
 XX PR 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX PA (LEUV-) LEUVEN RES & DEV.  
 XX PI Delcours J, Debysier W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 XX WP1; 2002-114579/15.  
 DR N-PSDB; ABK13674.  
 XX PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes

PS Claim 127; Page 57; 127pp; English.

XX The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable

CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial wheat TDXI-I (T. durum L endoxylanase  
 CC inhibitor).

XX SQ Sequence 196 AA;

Query Match 100.0%; Score 185; DB 23; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 6; Le-20;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAPVARAEAVAPFGVCYDRTKLTGNNLGGYAVPNV 35  
 |||||||||||||||||||||||||||||||||||

DB 81 GAPVARAEAVAPFGVCYDRTKLTGNNLGGYAVPNV 115

RESULT 5  
 AAU75815  
 ID AAU75815 standard; Protein: 370 AA.  
 XX AC AAU75815;  
 XX DT 23-APR-2002 (first entry)  
 XX DE Wheat L endoxylanase inhibitor, TAXI I, partial sequence TAXI-I.0.  
 XX KW Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX OS Triticum aestivum cultivar Soissons.  
 XX PN MO200198474-A1.  
 XX PD 27-DEC-2001.  
 XX PF 21-JUN-2001; 2001WO-BE00106.  
 XX PR 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX PA (LEUV-) LEUVEN RES & DEV.  
 XX PI Delcours J, Debysier W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 XX WP1; 2002-114579/15.  
 DR N-PSDB; ABK13672.  
 XX PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes

PS Claim 127; Page 57; 127pp; English.

XX The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the



KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX Triticum aestivum.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 95 /Label= unknown  
 FT Misc-difference 98 /Label= unknown  
 FT Misc-difference 101 /Label= unknown  
 FT Misc-difference 110 /Label= unknown  
 FT Misc-difference 333 /Label= unknown  
 FT Misc-difference /Label= unknown  
 XX  
 PN WO200198474-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 21-JUN-2001; 2001WO-BE00106.  
 XX  
 PR 22-JUN-2000; 2000GB-0015296.  
 PR 23-JAN-2001; 2001GB-0002018.  
 PR 16-MAR-2001; 2001GB-0002194.  
 PR 21-MAY-2001; 2001GB-0006564.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.  
 PI Delcours J, Debysers W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR WPI: 2002-114579/15.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes -  
 XX  
 PS Claim 127; Page 50-51; 127pp; English.  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, maltose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing, wheat gluten  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a variant of wheat TAXI-I (T. aestivum L. endoxylanase

CC inhibitor).  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.  
 XX  
 SQ Sequence 381 AA:  
 Query Match 97.3%; Score 180; DB 23; Length 381;  
 Best Local Similarity 97.1%; Pred. No. 8e-19;  
 Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GAPVARAVEAVAPFCGYCTKTLGNLGGYAVPNV 35  
 Db 266 GAPVARAVEVAPFVGVCYDTKTLGNLGGYAVPNV 300  
 RESULT 8  
 AAU75813  
 ID AAU75813 standard; Protein; 381 AA.  
 AC AAU75813;  
 XX  
 DT 23-APR-2002 (first entry)  
 DE Wheat L endoxylanase inhibitor, TAXI I, variant #2.  
 XX  
 KW Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Triticum aestivum.  
 FH  
 FT Key Location/Qualifiers  
 FT Misc-difference 95 /Label= unknown  
 FT Misc-difference 98 /Label= unknown  
 FT Misc-difference 101 /Label= unknown  
 FT Misc-difference 110 /Label= unknown  
 FT Misc-difference 333 /Label= unknown  
 FT Misc-difference /Label= unknown  
 XX  
 PN WO200198474-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 21-JUN-2001; 2001WO-BE00106.  
 XX  
 PR 22-JUN-2000; 2000GB-0015296.  
 PR 23-JAN-2001; 2001GB-0002018.  
 PR 16-MAR-2001; 2001GB-0002194.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.  
 PI Delcours J, Debysers W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR WPI: 2002-114579/15.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes -  
 XX  
 PS Claim 127; Page 51; 127pp; English.  
 CC The invention relates to separating and/or isolating inhibitors of



XX Sequence 381 AA;  
SQ Query Match 91.9%; Score 170; DB 23; Length 381;  
Best Local Similarity 94.3%; Pred. No. 2.7e-17;  
Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GAPVAAVEAVAPFGVCYDTKLTGNNLGGYAVPNV 35  
Db 266 GAPVAAVEAVAPFGVXXDTKLTGNNLGGYAVPNV 300  
RESULT 10  
AAU75824  
ID AAU75824 standard; Protein: 185 AA.  
XX AAU75824;  
AC  
XX 23-APR-2002 (first entry)  
XX  
XX Barley L endoxylanase inhibitor, HVXI I, variant #1.  
DE  
XX  
XX Barley: HVXI-I, L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
KW noodle; animal feed; starch separation; maize processing; malting;  
KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
XX  
XX Hordeum vulgare.  
OS  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 72 /label= Unknown  
FT Misc-difference 185 /label= Unknown  
FT  
XX  
XX WO200198474-A1.  
XX  
XX 27-DEC-2001.  
XX  
XX 21-JUN-2001; 2001WO-BE00106.  
XX  
XX 22-JUN-2000; 2000GB-0015296.  
XX 25-JAN-2001; 2001GB-0002018.  
XX 26-JAN-2001; 2001GB-0002194.  
XX 16-MAR-2001; 2001GB-0006564.  
XX 21-MAY-2001; 2001GB-0012328.  
XX  
XX (LEUV-) LEUVEN RES & DEV.  
XX  
XX Delcours J, Debysse W, Gebuere K, Goesaert H, Flerens K, Robben J;  
PI Van Campenhout S;  
XX  
XX WPI: 2002-114579/15.  
XX  
XX Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
PT beta-glucanolytic enzymes comprises using endoxylanases during  
PT screening for inhibition activity or affinity chromatography with  
PT immobilised enzymes  
XX  
XX Claim 127; Page 63; 127pp; English.  
XX  
XX The invention relates to separating and/or isolating inhibitors of  
XX cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
XX screening the inhibition activity by using two or more enzymes during the  
XX separation and/or isolation steps that allow to distinguish inhibitors of  
XX different specificity or by using an affinity chromatographic step with  
XX immobilised enzymes and/or antibodies against inhibitors. Also  
XX included are an isolated nucleic acid molecule encoding an inhibitor  
XX which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
XX alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
XX arabinoxylan or beta-glucan degrading enzymes, a vector comprising the

CC nucleic acid, an expression system transformed with the nucleic  
CC acid, a host organism transformed with the nucleic acid, the inhibitory  
CC proteins encoded by the nucleic acids and modulators of the proteins.  
CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
CC plant or plant materials transformed with the nucleic acid are useful  
CC for the formation of an endoxylanase-inhibitor complex, screening  
CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
CC reducing syruping in refrigerated dough compositions, affecting the  
CC relative affinity and/or relative hydrolysis specificity and/or relative  
CC hydrolysis rate versus water-extractable and/or water-unextractable  
CC arabinoxylans of endoxylanases such as by the formation of an  
CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
CC barley, sorghum and wheat and/or the production of beer, improving the  
CC production and/or quality of baked or extruded cereal products such as  
CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
CC biscuits, pasta and noodles, animal feed stuff, improving the production  
CC of starch separation and production, improving maize processing,  
CC plant disease resistance and nutraceutical and/or pharmaceutical  
CC applications, improving paper and pulp technologies. The present  
CC sequence is a variant of barley HVXI-I (H. vulgare L endoxylanase  
CC inhibitor).  
CC Note: Variant amino acids are highlighted in the specification but  
CC no wild-type sequence is shown for comparison.  
XX  
XX SQ Sequence 185 AA;  
XX  
XX Query Match 87.6%; Score 162; DB 23; Length 185;  
XX Best Local Similarity 82.9%; Pred. No. 1.8e-16;  
XX Matches 29; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GAPVAAVEAVAPFGVCYDTKLTGNNLGGYAVPNV 35  
Db 126 GAPVAAVEAVAPFGVXXDTKLTGNNLGGYAVPNV 160  
RESULT 11  
AA93763  
ID AA93763 standard; peptide; 57 AA.  
XX  
XX AA93763;  
XX  
XX 03-OCT-2000 (first entry)  
XX  
XX Amino acid sequence derived from an endo-beta-1,4-xylanase inhibitor.  
XX  
XX Endo-beta-1,4-xylanase inhibitor; wheat flour; xylanase; bakery product;  
XX dough; dough preparation.  
XX  
XX Triticum sp.  
XX  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 17 /note= "unspecified amino acid"  
FT Misc-difference 43 /note= "unspecified amino acid"  
FT Misc-difference 49 /note= "unspecified amino acid"  
FT  
XX  
XX WO200039289-A2.  
XX  
XX 06-JUL-2000.  
XX  
XX 17-DEC-1999; 99WO-IB02071.  
XX  
XX 23-DEC-1998; 98GB-0028599.  
XX 06-APR-1999; 99GB-0007805.  
XX 15-APR-1999; 99GB-0008645.  
XX  
XX (DANI-) DANISCO AS.  
XX  
XX Slibesen O, Sorensen JF;  
PI  
XX

DR	WPI: 2000-465744/40.
XX	Mutant xylanase protein identified using xylanase inhibitor useful for
PT	preparing non-sticky dough for bakery products .
XX	
PS	Disclosure: Page 106; 112pp; English.
XX	
CC	The present sequence is derived from an endo-beta-1,4-xylanase
CC	inhibitor. The protein is obtained from wheat flour. The specification
CC	also describes a mutant xylanase protein. The xylanase is useful for
CC	preparing a foodstuff, preferably a bakery product or a substance
CC	(e.g. a dough) for making the bakery product. Wild type xylanase or
CC	mutant xylanase is useful for preparing a dough that is less sticky
CC	than a dough comprising a fungal xylanase. The xylanase inhibitor is
CC	useful for screening high degree resistance xylinases for dough
CC	preparation. The xylanase is also useful for preparing a non-sticky
CC	dough. A combination of xylanase and the inhibitor is useful for
CC	calibrating and/or determining the quantity of inhibitor in a wheat
CC	flour sample.
XX	
SQ	Sequence 57 AA:
Query Match	78.4%; Score 145; DB 21; Length 57;
Best Local Similarity	87.9%; Pred. No. 1.6e-14;
Matches 29; Conservative	1; Mismatches 3; Indels 0; Gaps 0
OY	2 APVARAVEAVAFGYCYDPTKLTGNLGGVAPN 34           :           :           2 AVVARAVKDVAPEFGVXYDPKTGTGNNLGCVAPN 34
DJ	
RESULT 12	
AAU075826	
ID	AAU075826 standard; Protein: 185 AA.
XX	
AC	AAU075826;
DE	23-APR-2002 (first entry)
XX	
DE	Barley L endoxylanase inhibitor, HVXI I, microheterogenic variant.
KM	Barley: HVXI-I: L endoxylanase inhibitor; cellulolytic enzyme inhibitor; xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor; immobilised enzyme; enzyme; dough syringing; cereal product; beer; plant; straight dough; sponge dough; Chorleywood bread; biscuits; pasta; noodle; animal feed; starch separation; maize processing; malting; plant disease resistance; nutraceutical; pharmaceutical; paper; pulp. XX OS Hordeum vulgare. XX OS Synthetic.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 72 /label= Unknown
FT	Misc-difference 128 /label= Pro, Ala, Asp, Leu, Gly, Asn, Cys
FT	Misc-difference 132 /label= Pro, Ala, Asp, Leu, Gly, Asn, Cys
FT	Misc-difference 134 /label= Pro, Ala, Asp, Leu, Gly, Asn, Cys
FT	Misc-difference 148 /label= Pro, Ala, Asp, Leu, Gly, Asn, Cys
FT	Misc-difference 185 /label= Pro, Ala, Asp, Leu, Gly, Asn, Cys
XX	/label= unknown
PX	
PN	MO200198474-AI.
XX	
PD	27-DEC-2001.
PF	21-JUN-2001; 2001WO-BE00106.
XX	
XZ	22-JUN-2000; 2000GB-0015296.
PR	25-JAN-2001; 2001GB-0002018.

[illegible]

XX Barley; HMXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX Hordeum vulgare.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 72 /Label= unknown  
 FT Misc-difference 185 /Label= unknown  
 FT Misc-difference 185 /Label= unknown  
 PN WO200198474-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PD 21-JUN-2001; 2001WO-BE00106.  
 PF  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.  
 XX  
 PI Delcours J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 PI WPI: 2002-114579/15.  
 DR  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127: Page 63; 127pp: English.  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, xylan and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a variant of Barley HMXI-I (H. vulgare L endoxylanase

CC inhibitor).  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.  
 XX  
 SQ Sequence 185 AA:  
 Query Match 75.1%; Score 139; DB 23; Length 185;  
 Best Local Similarity 74.3%; Pred. No. 5.7e-13;  
 Matches 26; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 GAPVAAVEAVAPFGVCYDTKTLGNNGSYAVPNV 35  
 DB 126 GALAAGVNPVAPFGICYDAKTGNNGSYAVPNV 160  
 RESULT 14  
 AAU75820  
 ID AAU75820 standard; Protein; 102 AA.  
 XX  
 AC AAU75820;  
 XX  
 DT 23-APR-2002 (first entry)  
 DE  
 XX Rye L endoxylanase inhibitor, partial sequence SCXI-01.  
 DE  
 KW Rye; SCXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Secale cereale cultivar Halo.  
 XX  
 PN WO200198474-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PD 21-JUN-2001; 2001WO-BE00106.  
 PF  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.  
 XX  
 PI Delcours J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 PI WPI: 2002-114579/15.  
 DR N-PSDB; ARK13677.  
 DR  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127: Page 59; 127pp: English.  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory

protein encoded by the nucleic acids and modulators of the proteins.  
A recombinant protein, glycoprotein or polypeptide or microorganisms,  
plant or plant materials transformed with the nucleic acid are useful  
for the formation of an endoxylanase-inhibitor complex, screening  
endoxylanases that are totally, less or not inhibited by the inhibitors,  
reducing syringing in refrigerated dough compositions, affecting the  
relative affinity and/or relative hydrolysis specificity and/or relative  
hydrolysis rate versus water-extractable and/or water-unextractable  
arabinoxylans of endoxylanases such as by the formation of an  
endoxylanase/inhibitor complex, improving the malting of cereals such as  
barley, sorghum and wheat and/or the production of beer, improving the  
production and/or quality of baked or extruded cereal products such as  
straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
biscuits, pasta and noodles, animal feed stuff, improving the production  
of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
starch separation and production, improving maize processing,  
plant disease resistance and production, improving pharmaceutical  
applications, improving paper and pulp technologies. The present  
sequence represents a partial rye SXI-1 (S. cereale L endoxylanase  
inhibitor).

Sequence 102 AA;

Query Match	60.0%	Score 111	DB 23	Length 102
Best Local Similarity	91.3%	Pred. No. 5e-03		
Matches 21; Conservative	1	Mismatches	0	Gaps 0

```
Qy      1 GAPVARAVEA VAFGVCYDTKTL 23
          |||||: |||||
Db      80 GAPVARAVQ PVAPEGVCYDTKTL 102
```

RESULT 15  
AAU75821

ID AAU75821 standard; Protein; 102 AA

AC AAU75821;

DT 23-APR-2002 (first entry)

DE Rye L endoxylanase inhibitor, partial sequence SCXI-02

KW Rye: SCAR1: L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW xylinolitic enzyme inhibitor; dough stryping; cereal product; beer;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

OS Secale cereale cultivar Halo

PN WO200198474-A1.

PD 27-DEC-2001

PF 21-JUN-2001; 2001WO-BE00106

PR	22-JUN-2000;	2000GB-0015296
PR	25-JAN-2001;	2001GB-0002018
PR	26-JAN-2001;	2001GB-0002194
PR	16-MAR-2001;	2001GB-0006564
PR	21-MAY-2001;	2001GB-0012328

PA (LEUV-) LEUVEN RES &amp; DEV

PI Delcour J, Debyser W, Gebruers K, Goesaert H, Fierens K, Robben J,  
PI Van Campenhout S;

DR WPI; 2002-114579/15.  
DR N-PSDB; ABK13678.

PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or beta-glucanolytic enzymes comprising using endoxylanases during

PT screening for inhibition activity or affinity chromatography with  
PT immobilised enzymes -  
XX  
PS Claim 127; Page 60; 127pp; English.

The invention relates to separating and/or isolating inhibitors of cellulytic, xylanolytic and/or beta-glucanolytic enzymes comprises screening the inhibition activity by using two or more enzymes during the separation and/or isolation steps that allow to distinguish inhibitors of different specificity or by using an affinity chromatographic step with immobilised enzymes and/or antibodies against inhibitors. Also included are an isolated nucleic acid molecule encoding an inhibitor which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase, alpha-L-arabino-furanosidase, endoxylanase, and/or other cellulose, xylan, arabinoxylan or beta-glucan degrading enzymes, a vector comprising the nucleic acid, an expression system transformed with the nucleic acid, a host organism transformed with the nucleic acid, the inhibitory proteins encoded by the nucleic acids and modulators of the proteins. A recombinant protein, glycoprotein or polypeptide or microorganisms, plant or plant materials transformed with the nucleic acid are useful for the formation of an endoxylanase-inhibitor complex, screening endoxylanases that are totally, less or not inhibited by the inhibitors, reducing synerging in refrigerated dough compositions, affecting the relative affinity and/or relative hydrolysis specificity and/or relative hydrolysis rate versus water-extractable and/or water-unextractable arabinoxylans of endoxylanases such as by the formation of an endoxylanase/inhibitor complex, improving the malting of cereals such as barley, sorghum and wheat and/or the production of beer, improving the production and/or quality of baked or extruded cereal products such as straight dough, sponge dough, Chorlewood bread, breakfast cereals, biscuits, pasta and noodles, animal feed stuff, improving the production of starch derived syrups, sorbitol, xylitol and/or xylitol, wheat gluten starch separation and production, improving maize processing, plant disease resistance and production, improving and/or pharmaceutical applications, improving paper and pulp technologies. The present sequence represents a partial rye SCAR-I (S. cereale L endoxylanase inhibitor).

Sequence 102 AA;

Query Match	56.8%	Score 105:	DB 23:	Length 102:
Best Local Similarity	87.0%	Pred No. 4.1e-08:		
Matches 20: Conservative	1:	Mismatches 2:	Indels 0:	Gaps 0:

```
QY      1 GAPVARAVEA VAFGVCYDTKTL 23
          |||||:| |||||
Db      80 GAPVARAVKPV EFGVCYDTKTL 102
```

Search completed: May 20, 2003, 16:14:37  
Job time : 39.0769 secs



GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:11:22 ; Search time 14.2308 Seconds

(without alignments)  
236.439 Million cell updates/sec

Title: US-09-869-155-13

Perfect score: 185  
Sequence: 1 GAPVARAVEAVALPFGVCYDFTKLTGNNLGGYAVPNV 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79.5	43.0	433	2	F86163 hypothetical prote
2	79.5	43.0	434	2	G86163 hypothetical prote
3	62.5	33.8	448	1	S21426 conglutinin gamma pr
4	59	31.9	433	2	T14329 dermal glycoprotein
5	56.5	30.5	517	2	D82282 GMP synthase VC076
6	56.5	30.5	525	1	SYECGV GMP synthase (glut
7	56.5	30.5	525	2	AG1050 GMP synthetase [lm
8	56.5	30.5	525	2	AD0820 GMP synthase (glut
9	55.5	30.0	195	2	B81880 conserved hypothet
10	54.5	29.5	594	1	G81847 dihydrolipoamide d
11	54	29.2	427	1	S06750 basic 7S globulin
12	54	29.2	426	2	T02029 DNA-binding protei
13	53.5	28.9	311	2	B86793 dihydrocrotonate oxi
14	52.5	28.4	525	2	G83173 GMP synthase PA376
15	52	28.1	212	2	T05936 agglutinin isolect
16	52	28.1	242	2	F96717 hypothetical prote
17	52	28.1	340	2	T33394 probable glycerol
18	52	28.1	541	2	S01957 hypothetical prote
19	52	28.1	826	2	T46061 hypothetical prote
20	52	28.1	3206	1	GNVSPV genome polypeptide
21	51.5	27.8	478	2	E75359 glycolate oxidase
22	51.5	27.8	520	2	B95168 GMP synthase (limpo
23	51.5	27.8	526	2	C96034 GMP synthase (glut
24	51	27.6	556	2	A44441 B-cell antigen CD1
25	51	27.6	800	1	S31575 interleukin-4 rece
26	50.5	27.3	944	2	G90205 leucyl-tRNA synthe
27	50.5	27.3	304	2	A98146 probable threonin
28	50.5	27.3	336	2	AC3142 threonine dehydrat
29	50.5	27.3	495	2	H69614 aldehyde dehydroge

30	50.5	27.3	525	2	F85894 GMP synthetase (gl
31	50.5	27.3	525	2	G97392 gmp synthase (glut
32	50.5	27.3	525	2	A12610 GMP synthase (limpo
33	50	27.0	288	2	S73016 H+-transporting tw
34	50	27.0	430	2	T48142 B-lymphocyte antig
35	50	27.0	432	2	T32044 hypothetical prote
36	50	27.0	534	2	JC4572 signal recognition
37	49.5	26.8	195	2	T46854 anthranilate synth
38	49.5	26.8	360	2	D86200 protein F12K11.20
39	49.5	26.8	510	2	E70321 GMP synthase (glut
40	49.5	26.8	522	2	A82558 glutamine amidotra
41	49	26.5	300	1	JC5299 protein kinase (EC
42	49	26.5	444	2	B95033 aminopeptidase C I
43	49	26.5	444	2	B97904 aminopeptidase C I
44	49	26.5	538	2	E96492 hypothetical prote
45	49	26.5	594	1	S42920 dihydrolipoamide d

## ALIGNMENTS

RESULT 1  
F86163  
hypothetical protein F15K9.17 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #extl\_change 27-Nov-2001  
C:Accession: F86163  
R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: F86163  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-433 <STO>  
A:Cross-references: GB:AEO05172; NID:g3850579; PIDN:AAC72119.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
C:Superfamily: conglutinin gamma  
Query Match  
Best Local Similarity 43.0%; Score 79.5; DB 2; Length 433;  
Matches 16; Conservative 47.1%; Pred. No. 0.002;  
Mismatch 9; Indels 3; Gaps 1;  
QY 5 ARAVEA---PRGVCYDFTKLTGNNLGGYAVPNV 35  
II:::II IIII:II IIII:II IIII:II  
DB 315 ARSIRKAVSVKPGACFTKNNVTRIGYAVPEI 348  
RESULT 2  
F86163  
hypothetical protein F15K9.16 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #extl\_change 31-Dec-2001  
C:Accession: G86163  
R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712

f:Accession: G86163  
 f:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-434 <SPO>  
 A:Cross-references: GB:AE005172; NID:93850580; PIDN:AACT2120.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: conglutin gamma

Query Match 43.0%; Score 79.5; DB 2; Length 434;  
 Best Local Similarity 47.1%; Pred. No. 0.002;  
 Matches 16; Conservative 6; Mismatches 9; Indels 3; Gaps 1;

QY 5 ARAVEAVA---PGVCYDTKTLGNLGGYAVPNV 35  
 DB 316 ARSIRKVASVKPGACFSTKNVGTSLGYAVPEI 349

## RESULT 3

S21426

conglutin gamma precursor - narrow-leaved blue lupine

C:Species: Lupinus angustifolius (narrow-leaved blue lupine)

C&gt;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S21426; S31092

R:Koliyas, S.; Gayler, K.R.  
 Submitted to the EMBL Data Library, April 1992

A:Reference number: S21426

A:Accession: S21426

A:Molecule type: mRNA

A:Residues: 1-448 &lt;KOL1&gt;

A:Cross-references: EMBL:X65601

A:Note: It is uncertain whether Met-1 or Met-5 is the initiator

R:Koliyas, S.; Gayler, K.R.  
 Plant Mol. Biol. 21, 397-401, 1993

A&gt;Title: Structure of the cDNA coding for conglutin gamma, a sulphur-rich protein from L

A:Reference number: S31092; NMID:93144715; PMID:8425065

A:Accession: S31092

A:Molecule type: mRNA

A:Residues: 5-448 &lt;KOL2&gt;

A:Cross-references: EMBL:X65601

C:Superfamily: conglutin gamma

C:Keywords: disulfide bond; seed; storage protein

F:..33/Domain: signal sequence #status predicted &lt;SIG&gt;

F:..294/Product: conglutin gamma large chain #status predicted &lt;MA1&gt;

F:..95-448/Product: conglutin gamma small chain #status predicted &lt;MA2&gt;

Query Match 33.8%; Score 62.5; DB 1; Length 448;  
 Best Local Similarity 42.4%; Pred. No. 0.57;  
 Matches 14; Conservative 6; Mismatches 8; Indels 5; Gaps 2;

QY 3 PVARAVEAVAPFGVCYDTKTLGNLGGYAVPNV 35  
 DB 335 PKQAVKAVGPGFLCYDSRKIS--GG-APSV 362

## RESULT 4

T14329

dermal glycoprotein precursor, extracellular - carrot (fragment)

C:Species: Daucus carota (carrot)

C&gt;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Nov-2000

C:Accession: T14329

R:Stitch, S.; Sturm, A.; Fujii, T.; Chrispeels, M.J.  
 Planta 188, 432-438, 1992

A&gt;Title: cDNA cloning of an extracellular dermal glycoprotein of carrot and its expressi

A:Reference number: Z17987

A:Accession: T14329

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-433 &lt;SAT&gt;

A:Cross-references: EMBL:D14550; NID:d1004922; PID:d1003923

A:Experimental source: subspecies US-Harumakigosun; root

C:Genetics:  
 A:Gene: EDGP1

C:Superfamily: conglutin gamma  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-433/Product: dermal glycoprotein, extracellular #status predicted <MAT>

Query Match 31.9%; Score 59; DB 2; Length 433;  
 Best Local Similarity 39.3%; Pred. No. 1.7;  
 Matches 11; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 8 VEAVAPFGVCYDTKTLGNLGGYAVPNV 35  
 DB 323 VASVAPPGACFSTNDILSTRLGPSPSI 350

## RESULT 5

D82282

GMP synthase VC0768 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C&gt;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: D82282

R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.  
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000

A&gt;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; NMID:20406833; PMID:10952301

A:Accession: D82282

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-517 &lt;HEI&gt;

A:Cross-references: GB:AE004162; GB:AE003852; NID:9655211; PIDN:AAF93933.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0768

A:Map position: 1

C:Superfamily: GMP synthase (glutamine-hydrolyzing); trpC homology

Query Match 30.5%; Score 56.5; DB 2; Length 517;  
 Best Local Similarity 41.4%; Pred. No. 4.8;  
 Matches 12; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

QY 1 GAPVAREAVAPFGVCYDTKTLGNLGG 29  
 DB 79 GVPV-----FGVCYGMOTMAEQLGG 98

## RESULT 6

SYECGU

GMP synthase (glutamine-hydrolyzing) (EC 6.3.5.2) [validated] - Escherichia coli (str

C:Species: Escherichia coli

C&gt;Date: 17-Mar-1987 #sequence\_revision 30-Sep-1987 #text\_change 01-Mar-2002

C:Accession: A24640; B65027; A01197

R:riedeman, A.A.; Smith, J.M.; Zalkin, H.  
 J. Biol. Chem. 260, 8676-8679, 1985

A&gt;Title: Nucleotide sequence of the gua gene encoding GMP synthetase of Escherichia

A:Reference number: A92546; NMID:85261223; PMID:3694545

A:Accession: A24640

A:Molecule type: DNA

A:Residues: 1-525 &lt;TIE&gt;

A:Cross-references: GB:M10101; GB:M10102; NID:g146274; PIDN:AA818619.1; PID:g146276

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997

A&gt;Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; NMID:97426617; PMID:9278503

A:Accession: B65027

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-525 &lt;BLAT&gt;

A:Cross-references: GB:AE000337; GB:U00096; NID:q1788850; PIDN:AACT5560.1; PID:g17888

A:Experimental source: strain K-12, substrain MG1655

C:Comment: The active enzyme catalyzes the formation of GMP by transferring the amide

C:Genetics:  
 A:Gene: guaA



## RESULT 11

Basic 7S globulin precursor - soybean

N:Alternate names: Insulin-binding protein

C:Species: Glycine max (soybean)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S06750; S00218; A61003; JN0311; JN0312

R:Kagawa, H.; Hirano, H.

Nucleic Acids Res. 17, 8868, 1989

A:Title: Sequence of a cDNA encoding soybean basic 7S globulin.

A:Reference number: S06750; M0ID:90067863; PMID:2587227

A:Accession: S06750

A:Molecule type: mRNA

A:Residues: 1-427 <KAI>

A:Cross-references: EMBL:X16469; NID:q18542; PIDN:CAA34489.1; PID:q18543

R:Kagawa, H.; Yamauchi, F.; Hirano, H.

FEBS Lett. 226, 145-149, 1987

A:Title: Soybean basic 7S globulin represents a protein widely distributed in legume spe

A:Reference number: S00218

A:Accession: S00218

A:Molecule type: protein

A:Residues: 276-293, 'X'; 295-304, 'FX'; 423-427, 'X'; 'IVGPFGLCPNNGVTSGLPMXXMPAR'; 'QL', 356,

42, 'AX' <KAG>

R:Hirano, H.; Matanabe, T.

Electrophoresis 11, 573-580, 1990

A:Title: Microsequencing of proteins electrophoretically transferred onto immobilizing matrices from

A:Reference number: A61003; M0ID:91031409; PMID:2226413

A:Accession: A61003

A:Molecule type: protein

A:Residues: 25-57;135-150;276-304, 'X', 306-308, 'X', 310-311;383-417 <HIR>

R:Barabshov, S.F.; Egorov, T.A.; Kochkina, V.M.

Bioorg. Khim. 17, 421-423, 1991

A:Title: Isolation and characterization of soybean insulin-binding protein.

A:Reference number: JN0309; M0ID:91291181; PMID:2064630

A:Accession: JN0311

A:Molecule type: protein

A:Residues: 'VPVPOHKTN', 27-33, 'X', 35-42, 'XX', 45, 'X', 47, 'X', 49, 'AX', 52 <BAR>

A:Note: 20-Thr, 21-Asn, 25-Pro and 50-Asn were also found

A:Accession: JN0312

A:Molecule type: protein

A:Residues: 'S', 99, 'IV', 102, 'SX', 105, 'C', 286-299, 'XX', 302, 'XX', 305 <BA2>

A:Note: 102-Gly and 296-Lys were also found.

C:Comment: This protein specifically binds to bovine insulin.

C:Superfamily: conglutin gamma

C:Keywords: heterodimer; seed; storage protein

3;1-24/Domain: signal sequence #status predicted <SIG>

3;23-27/Product: basic 7S globulin heavy chain #status experimental <MAT1>

1;276-427/Product: basic 7S globulin light chain #status experimental <MAT2>

Query Match 29.2%; Score 54; DB 1; Length 427;

Best Local Similarity 36.4%; Pred. No. 8.9;

Matches 12; Conservative 8; Mismatches 7; Indels 6; Gaps 2;

OY 3 PVARAVEAVAFEGVCYDPTKTLGNLGGYAVPNV 35

DB 314 PKQAVKSVAPFGLCFNS---NKINAY--PSV 340

## RESULT 12

T03029 DNA-binding protein pabf - common tobacco

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 14-May-1999

C:Accession: T02029; T02342

R:Jaidle, G.; Natt, E.; Seguin, A.; Doerner, P.W.; Lamb, C.J.

submitted to the EMBL Data Library, February 1994

A:Description: A novel DNA-binding protein with histone H1 and high mobility group prote

rometer.

A:Reference number: Z14501

A:Accession: T02029

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-546 <LAI>

A:Cross-references: EMBL:U06712; NID:9555654; PID:9555655

A:Experimental source: strain Xanthi; tissue-type stem

R:Tjden, G.; Coruzzi, G.M.

Plant Cell 1, 107-118, 1994

A:Title: A novel At-rich DNA binding protein that combines a putative transcription

A:Reference number: Z14662

A:Accession: T02342

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 167-390, 'G', 392-546 <TJA>

A:Cross-references: EMBL:I26113; NID:9456123; PID:9456124

A:Experimental source: cultivar SRI; tissue-type leaf

C:Genetics:

A:Gene: pabf; ATPB-1

C:Keywords: DNA binding

Query Match 29.2%; Score 54; DB 2; Length 546;

Best Local Similarity 35.3%; Pred. No. 11;

Matches 12; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

OY 1 GAPVAVEAVAFEGVCYDPTKTLGNLGGYAVPN 34

DB 400 GPAAATVGVTDVPIAAFTDNLPAVGGGVYN 433

## RESULT 13

dihydroorotate oxidase (EC 1.3.3.1) [imported] - Lactococcus lactis subsp. lactis (st

B86793

N:Alternate names: dihydroorotate dehydrogenase B

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001

C:Accession: B86793

R:Bodlwin, A.; Muncher, P.; Manger, S.; Jallion, O.; Malarne, K.; Weissenbach, J.; Eh

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis

A:Reference number: A86625; M0ID:21235186; PMID:11337471

A:Accession: B86793

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-311 <STO>

A:Cross-references: GB:AE005176; PID:q12724328; PIDN:AAK05444.1; GSPDB:GM00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: pydB

C:Superfamily: dihydroorotate oxidase

C:Keywords: oxidoreductase

Query Match 28.9%; Score 53.5; DB 2; Length 311;

Best Local Similarity 34.8%; Pred. No. 7.5;

Matches 16; Conservative 5; Mismatches 12; Indels 13; Gaps 2;

OY 3 PVARAVEAVAFEGVCYDPTKTLGNLGGYAVPNV 35

DB 180 PIRAAVEAGACDGLTINTLMGVRFDLTKRPVLANITGSLSPAI 225

## RESULT 14

GMP synthase PA3769 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: G83173

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

.; Lory, S.; Olson, M.V.

Nature 400, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; M0ID:20437337; PMID:10984043

A:Accession: G83173

A:Status: preliminary

A:Molecule type: DNA



**THIS PAGE BLANK (uspto)**

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:05:42 : Search time 9.42308 Seconds

(without alignments)  
154.055 million cell updates/sec

Title: US-09-869-155-13

Sequence: 1 GAPVRAVEAVAPFGVCYDTKLTGNLGYAVPNV 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56.5	30.5	517	1 GUA_VIBCH	09ktw2 vibrio chol
2	56.5	30.5	525	1 GUA_ECOLI	P04079 escherichia
3	54	29.2	427	1 7SBG_SOYBN	P13917 glycine max
4	53.5	28.9	311	1 PYDB_IACLA	09ctw8 lactococcus
5	53.5	28.9	311	1 PYDB_IACLC	P54322 lactococcus
6	52.5	28.4	525	1 GUA_PSEAE	09hxm6 pseudomonas
7	52	28.1	212	1 AGI_HORVU	P15312 hordeum vul
8	52	28.1	3206	1 POLG_PSBMV	P29152 P genome po
9	51	27.6	556	1 CD19_HUMAN	P15391 homo sapien
10	51	27.6	944	1 SYL2_SULSO	033768 sulfobus
11	50.5	27.3	500	1 DHAL_AGABI	074187 agaricus bl
12	50.5	27.3	525	1 GUA_AGR75	08u112 agrobacteri
13	50	27.0	430	1 CD19_CAVPO	P25179 cavia porce
14	50	27.0	432	1 YK27_CWEL	P16686 caenorhabdl
15	50	27.0	534	1 SR54_ASPNG	000179 aspergillus
16	49.5	26.8	510	1 GUA_AQUAE	066601 aquifex aeo
17	49.5	26.8	522	1 GUA_XYLFA	09par6 xylicella fas
18	48.5	26.2	255	1 AMPH_TREPA	083614 treponema p
19	48.5	26.2	524	1 GUA_CORAM	052831 corynebacte
20	48.5	26.2	526	1 GUA_STRCO	0910h2 streptomyce
21	48	25.9	186	1 AG13_WHEAT	P10969 triticum ae
22	48	25.9	213	1 AG12_WHEAT	P02876 triticum ae
23	48	25.9	901	1 MSH5_YEAST	012175 saccharomyc
24	48	25.9	1004	1 SLPO_BACBR	P09333 bacillus br
25	48	25.9	1608	1 HLTA_SERMA	P15320 serratia ma
26	47.5	25.7	2185	1 PYRI_DICDI	P20054 dictyostell
27	47.5	25.7	335	1 GPD_A_PSELE	059680 pseudomonas
28	47.5	25.7	523	1 GUA_AHAEN	P44335 haemophilus
29	47	25.4	331	1 YC65_ARCFU	029003 archaeoglob
30	47	25.4	201	1 MCP_BPR41	P26596 lactococcus
31	47	25.4	432	1 PROA_DEIRA	09rtcd9 delnoccocus
32	47	25.4	477	1 NIFE_RHOCA	P19055 rhodobacter
33	47	25.4	547	1 CD19_MOUSE	P25918 mus musculu

34	47	25.4	675	1 CAA_BACTI	P09662 bacillus th
35	47	25.4	915	1 SMOO_HUMAN	P53814 homo sapien
36	46.5	25.1	521	1 GUA_METMB	09ixr2 neisseria m
37	46.5	25.1	589	1 GLMS_METTH	026273 m glucosami
38	46.5	25.1	782	1 DPO2_ECOLI	P21189 escherichia
39	46.5	25.1	4485	1 DYHG_CHLRE	039575 chlamydomon
40	46	24.9	269	1 MURI_PASMO	P57887 pasteurella
41	46	24.9	430	1 GSA_BACSU	P30949 bacillus su
42	46	24.9	432	1 GSA_VIBCH	09ku97 vibrio chol
43	46	24.9	507	1 GUA_BRUME	08y112 bruceella me
44	46	24.9	571	1 PRDE_HUMAN	09gzv8 homo sapien
45	46	24.9	614	1 BTUB_ECOLI	P06129 escherichia

## ALIGNMENTS

```

RESULT 1
ID      GUA_VIBCH      STANDARD:      PRT:      517 AA.
AC      09KTW2:
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2) (Glutamine
DE      amidotransferase) (GMP synthetase).
GN      GUA OR VC0768.
OS      Vibrio cholerae.
OC      Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX      NCBI_TaxID=666;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=El Tor N16961 / Serotype O1;
RX      MEDLINE=20406833; PubMed=10952301;
RA      Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Ginn M.L.,
RA      Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA      Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA      Esmolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA      McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,
RA      Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA      Fraser C.M.;
RT      "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT      cholerae.";
RL      Nature 406:477-483(2000).
CC      -1- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
CC      H(2O) -> AMP + diphosphate + GMP + L-glutamate.
CC      -1- PATHWAY: GMP biosynthesis.
CC      -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC      -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMP SYNTHASE
CC      FAMILY.
CC      -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AE004162; AAF93933.1; -
CC      HSSP: P04079; IGPM.
CC      TIGR: VC0768; -
DR      InterPro: IPR002385; AntH_synthIII.
DR      InterPro: IPR001317; CPS_GATase.
DR      InterPro: IPR000991; GATase_1.
DR      InterPro: IPR001674; GMP_synth_C.
DR      InterPro: IPR004739; GMPsynthase_N.
DR      Pfam: PF00958; GMP_synth_C_1.
DR      PRINTS: PR00097; ANTSYNTHASEII.
DR      PRINTS: PR00099; CPSGATASE.
DR      PRINTS: PR00096; GATASE.

```

```

DR TIGR00884; guaa_Cterm; 1.
DR TIGR00888; guaa_Nterm; 1.
DR PROSITE; PS00442; GATASE_TYPE.1; 1.
KW Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
KM Glutamine amidotransferase; Complete proteome.
FT DOMAIN 1 198
FT ACT_SITE 231 393
FT ACT_SITE 86 86
FT ACT_SITE 173 173
FT ACT_SITE 175 175
FT NE_BIND 227 233
SQ SEQUENCE 517 AA; 57764 MW; 80AC2194E1172C45 CRC64;

Query Match 30.5%; Score 56.5; DB 1; Length 517;
Best Local Similarity 41.4%; Pred. No. 3.1;
Matches 12; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

QY 1 GAPPARAVEVAPPGVCYDRTKGNLNGC 29
Db 79 GVPV-----FGVCYGMQTMALQGLG 98

RESULT 2
GUA_ECOLI STANDARD; PRT; 525 AA.
ID GUA_ECOLI
AC P04079;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2) (glutamine
DE amidotransferase) (GMP synthetase).
GN GUA OR B2507.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
FN [1]
FP SEQUENCE FROM N.A.
FX MEDLINE=8526123; PubMed=3894345;
RA Tiedeman A.A., Smith J.M., Zalkin H.;
RT "Nucleotide sequence of the guaa gene encoding GMP synthetase of
RT Escherichia coli K12.";
RL J. Biol. Chem. 260:8676-8679(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / NG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubraman S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [4]
RP GATASE DOMAIN.
RX MEDLINE=85131126; PubMed=2982857;
RA Zalkin H., Argos P., Narayana S.V.L., Tiedeman A.A., Smith J.M.;
RT "Identification of a trpG-related glutamine amide transfer domain in
RT Escherichia coli GMP synthetase.";
```

```

RL J. Biol. Chem. 260:3350-3354(1985).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE=96133732; PubMed=8548458;
RA Tesmer J.J.G., Klem T.J., Deras M.L., Davison V.J., Smith J.L.;
RT "The crystal structure of GMP synthetase reveals a novel catalytic
RT triad and is a structural paradigm for two enzyme families.";
RL Nat. Struct. Biol. 3:74-86(1996).
CC -1- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
CC H(2)O = AMP + diphosphate + GMP + L-glutamate.
CC -1- PATHWAY: GMP biosynthesis.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMP SYNTHASE
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M10101; AAB18619.1; -.
CC DR EMBL: AE000337; AAC75560.1; -.
CC DR EMBL: D90880; BAA16394.1; -.
CC DR PIR: A26440; SYECGU.
CC DR PDB: 1GPM; 29-JAN-96.
CC Ecogen: EG10420; guaa.
CC DR InterPro: IPR000991; GATase_1.
CC DR InterPro: IPR001674; GMP_synth_C.
CC DR InterPro: IPR004739; GMPsynthase_N.
CC DR InterPro: IPR004135; tRNA_Me_trans.
CC DR Pfam: PF00117; GATase; 1.
CC DR Pfam: PF00958; GMP_synth_C; 1.
CC DR Pfam: PF03054; tRNA_Me_trans; 1.
CC DR PRINTS: PR00096; GATASE.
CC DR TIGR00884; guaa_Cterm; 1.
CC DR TIGR00888; guaa_Nterm; 1.
CC DR PROSITE: PS00442; GATASE_TYPE.1; 1.
KW Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
KW Glutamine amidotransferase; 3D-structure; Complete proteome.
FT DOMAIN 1 206
FT FT 239 401
FT ACT_SITE 86 86
FT ACT_SITE 181 181
FT ACT_SITE 183 183
FT NE_BIND 235 241
SQ SEQUENCE 525 AA; 58679 MW; D934786DF3D694B CRC64;

Query Match 30.5%; Score 56.5; DB 1; Length 525;
Best Local Similarity 52.2%; Pred. No. 3.2;
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 9 EAVAP-FCVCYDRTKGNLNGC 30
Db 77 EACVPVFCVCYGMQTMALQGLG 99

RESULT 3
7SBG_SOYBN STANDARD; PRT; 427 AA.
ID 7SBG_SOYBN
AC P13917; Q39901; Q43464;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Basic 7S globulin precursor (Bg) (SBg7S).
GN BG OR G7S.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
```



```

OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=cv. Miyagishihime; TISSUE=Seed;
RX MEDLINE=90067863; PubMed=2587227;
RA Kagawa H., Hirano H.;
RT "Sequence of a cDNA encoding soybean basic 7S globulin.";
RN Nucleic Acids Res. 17:8868-8868(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Miyagishihime; TISSUE=Etolated leaf;
RX MEDLINE=94336768; PubMed=8058830;
RA Watanabe Y., Hirano H.;
RT "Nucleotide sequence of the basic 7S globulin gene from soybean.";
RN Plant Physiol. 105:1019-1020(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Shi-shi; TISSUE=Cotyledon;
RX Shu T.F., Hsieh K.L., Hsing Y.I., Chen Z.Y., Chow T.Y.;
RT Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE.
RA Kagawa H., Yamauchi F., Hirano H.;
RT "Soybean basic 7S globulin represents a protein widely distributed in
legume species.";
RN FEBS Lett. 226:145-149(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN. HAS A PROTEIN KINASE ACTIVITY.
CC BINDS LEGUMININ.
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE BG IS COMPOSED OF HIGH- AND LOW-KDA SUBUNITS
WHICH ARE LINKED BY DISULFIDE BONDING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X16469; CAA34489.1; -
DR EMBL: U59425; AAR03390.1; -
DR EMBL: D16107; BAA03681.1; -
DR PIR: S06750; S06750.
CC Glycoprotein; Seed storage protein; signal.
FT SIGNAL 1 24
FT CHAIN 25 427 BASIC 7S GLOBULIN.
FT CHAIN 25 275 HIGH-KDA SUBUNIT.
FT CHAIN 276 427 LOW-KDA SUBUNIT.
FT CONFLICT 48 48 W -> S (IN REF. 3).
FT CONFLICT 264 264 N -> T (IN REF. 1).
FT CONFLICT 305 305 F -> C (IN REF. 1).
FT CONFLICT 309 309 F -> C (IN REF. 1).
SQ SEQUENCE 427 AA; 46393 MW; 66041BC0680BACB CRC64;

Query Match 29.2%; Score 54; DB 1; Length 427;
Best Local Similarity 36.4%; Pred. No. 5.8;
Matches 12; Conservative 8; Mismatches 7; Indels 6; Gaps 2;

QY 3 PVARAVEAVAPFGVCYDTKTLGNLGGYAVPNV 35
DB 314 PROADQKVSAPFGLCFNS---NKINAV--PSV 340

RESULT 4
PYDB_LACLA STANDARD; PRT; 311 AA.
AC Q9CFW8;
DR 16-OCT-2001 (Rel. 40; Created)
DR 16-OCT-2001 (Rel. 40; Last sequence update)
DR 15-JUN-2002 (Rel. 41; Last annotation update)
DE Dihydroorotate dehydrogenase B, catalytic subunit (EC 1.3.3.1)
DE (Dihydroorotate oxidase B) (DHODase B) (DHOD B).

```

```

GN PYRDB OR PYDB OR LL1346.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jallion O., Malarme K.,
Welschenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403.";
RN Genome Res. 11:731-753(2001).
CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) = orotate +
H(2O(2)).
CC -1- COFACTOR: FMN (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; fourth step.
CC -1- SUBUNIT: HETEROETRAMER OF 2 PYR AND 2 PYRDB SUBUNITS (By
similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE DIHYDROOROTATE DEHYDROGENASE FAMILY.
SUBFAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE006366; AAK05444.1; -
DR HSSP: P54321; 2DOR.
DR InterPro: IPR001295; DHO_dh.
DR InterPro: IPR003009; FMN_enzyme.
DR Pfam: PF01180; DHODase; 1.
DR TIGRfam: TIGR01037; pyrd_sub1_fam; 1.
DR PROSITE: PS00911; DHODENASE_1; 1.
DR PROSITE: PS00912; DHODENASE_2; 1.
KW Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN;
Complete proteome.
FT NP_BIND 242 250 FMN (POTENTIAL).
FT SEQUENCE 311 AA; 32916 MW; 6FE419606B2E733C CRC64;

Query Match 28.9%; Score 53.5; DB 1; Length 311;
Best Local Similarity 34.8%; Pred. No. 5;
Matches 16; Conservative 5; Mismatches 12; Indels 13; Gaps 2;

QY 3 PVARAVEAVAP-----FGVCYDTKT---LGNLGGYAVPNV 35
DB 180 PIKAVEAGAGDGLTMINILMGVRFDLKRPVLANITGSLSPAI 225

RESULT 5
PYDB_LACLC STANDARD; PRT; 311 AA.
AC P54322;
DR 01-OCT-1996 (Rel. 34; Created)
DR 01-OCT-1996 (Rel. 34; Last sequence update)
DR 15-JUN-2002 (Rel. 41; Last annotation update)
DE Dihydroorotate dehydrogenase B, catalytic subunit (EC 1.3.3.1)
DE (Dihydroorotate oxidase B) (DHODase B) (DHOD B).
GN PYRDB.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MG1363;
RX MEDLINE=94292416; PubMed=8021180;
RA Andersen P.S., Jansen P.J.G., Hammer K.;
RT "Two different dihydroorotate dehydrogenases in Lactococcus lactis.";
RN J. Bacteriol. 176:3975-3982(1994).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MG1363;
RX MEDLINE-96326348; PubMed-8759867;
RA Andersen P.S., Martinussen J., Hammer K.;
RT "Sequence analysis and identification of the pyrKDF operon from
RT Lactococcus lactis including a novel gene, pyrK, involved in
RT pyrimidine biosynthesis."
RL J. Bacteriol. 178:5005-5012(1996).
RN [3]
RP CHARACTERIZATION.
RC STRAIN-MG1363;
RX MEDLINE-97067197; PubMed-8910559;
RA Nielsen F.S., Andersen P.S., Jensen K.F.;
RT "The B form of dihydroorotate dehydrogenase from Lactococcus lactis
RT consists of two different subunits, encoded by the pyrB and pyrK
RT genes, and contains FMN, FAD, and [Fes] redox centers."
RL J. Biol. Chem. 271:29359-29365(1996).
CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) -> orotate +
CC H(2)O(2).
CC -1- COFACTOR: FMN.
CC -1- PATHWAY: pyrimidine biosynthesis; fourth step.
CC -1- SUBUNIT: HETEROTRIMER OF 2 PYR AND 2 PYRB SUBUNITS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).
CC -1- MISCELLANEOUS: CAN USE NAD BUT NOT FUMARATE AS AN ELECTRON
CC ACCEPTOR.
CC -1- SIMILARITY: BELONGS TO THE DIHYDROOROTATE DEHYDROGENASE FAMILY.
CC SUBFAMILY 1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X74207; CAA52280.1; -
DR HSSP: P54321; 2DOR.
DR InterPro: IPR001295; DHO_dh.
DR InterPro: IPR003009; FMN_enzyme.
DR Pfam: PF01180; DHODEHase; 1.
DR TIGRPFAM: TIGR01037; pyrD_sub1_fam; 1.
DR PROSITE: PS00911; DHODEHase_1; 1.
DR PROSITE: PS00912; DHODEHase_2; 1.
DR Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN.
FT NP_BIND 242 250 FMN (POTENTIAL).
FT SEQUENCE 311 AA; 33100 MW; E64C5B635F57B6EE CRC64;
OY 3 PYRAVEAVAP-----FGVCYDRTK-----LGNNGSYAIPNV 35
Db 180 PIKAAVEAAGADGIMINILMGVRFDKTRPIILANITGSLSPAI 225

Query Match 28.9%; Score 53.5; DB 1; Length 311;
Best Local Similarity 34.8%; Pred. No. 5;
Matches 16; Conservative 5; Mismatches 12; Indels 13; Gaps 2;

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE-20437337; PubMed-10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Gardner R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
CC H(2)O = AMP + diphosphate + GMP + L-glutamate.
CC -1- PATHWAY: GMP biosynthesis.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO THE GMP SYNTHASE
CC FAMILY.
-----
CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AE004796; AAC07156.1; -
DR HSSP: P04079; 1GPM.
DR InterPro: IPR002385; Anth_synthII.
DR InterPro: IPR001317; CPS_GATase.
DR InterPro: IPR000991; GATase_1.
DR InterPro: IPR001674; GMP_synth_C.
DR InterPro: IPR004739; GMPsynthase_N.
DR Pfam: PF00117; GATase_1.
DR Pfam: PF00958; GMP_synth_C; 1.
DR PRINTS: PR00097; ANTSNTASEII.
DR PRINTS: PR00099; CPSGATASE.
DR PRINTS: PR00096; GATASE.
DR TIGRPFAM: TIGR00884; guaA_cleam; 1.
DR PROSITE: PS00442; GATASE_TYPE1; 1.
DR Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
KW Glutamine amidotransferase; Complete proteome.
FT DOMAIN 1 206 GLUTAMINE AMIDOTRANSFERASE.
FT DOMAIN 239 401 GMP-BINDING (BY SIMILARITY).
FT ACT_SITE 86 86 GATASE (BY SIMILARITY).
FT ACT_SITE 181 181 GATASE (BY SIMILARITY).
FT ACT_SITE 183 183 GATASE (BY SIMILARITY).
FT NP_BIND 235 241 ATP (BY SIMILARITY).
FT SEQUENCE 525 AA; 57959 MW; CDBELICE73C3C40 CRC64;
OY 6 RAVEAV-----AP-FGVCYDRTKGNLGG 29
Db 70 RAQAVFPLDKVPLFGICYGQMTAEQMG 98

Query Match 28.4%; Score 52.5; DB 1; Length 525;
Best Local Similarity 41.4%; Pred. No. 11;
Matches 12; Conservative 5; Mismatches 7; Indels 5; Gaps 2;

```

```

RESULT 6
GUA_PSEAE
ID GUA_PSEAE STANDARD: PRT; 525 AA.
AC Q9HXK6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2) (Glutamine
DE amidotransferase) (GMP synthetase).
GN GUA OR PA3769.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]

```

```

RESULT 7
AGI_HORVU
ID AGI_HORVU STANDARD: PRT; 212 AA.
AC P15312;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Root-specific lectin precursor.
DE Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;

```

OC Trilicaceae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lerner D.R., Raikhel N.V.;  
 RT "Cloning and characterization of root-specific barley lectin.";  
 RL Plant Physiol. 91:124-129(1989).  
 CC -i- FUNCTION: CARBOHYDRATE BINDING.  
 CC -i- SIMILARITY: CONTAINS 4 CHITIN-BINDING DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M29280; AAA32969.1; .  
 CC HSSP: P10969; 1MGT.  
 CC InterPro: IPR001002; Chitin\_binding\_1.  
 CC Pfam: PF00187; Chitin\_bind\_1; 4.  
 CC PRINTS: PR00451; CHITINBINDING.  
 CC ProDom: PD000609; Chitin\_binding\_1; 4.  
 CC SMART: SM00270; ChbD1; 4.  
 CC ProSITE: PS00026; CHITIN\_BINDING; 4.  
 CC Lectin: Repeat: Chitin-binding; Glycoprotein; Signal.  
 FT SIGNAL 1 26  
 FT CHAIN 27 212  
 FT DOMAIN 27 69 ROOT-SPECIFIC LECTIN.  
 FT DOMAIN 70 112 CHITIN-BINDING 1.  
 FT DOMAIN 113 155 CHITIN-BINDING 2.  
 FT DOMAIN 156 197 CHITIN-BINDING 3.  
 FT MOD\_RES 27 27 PYRROLIDONE CARBOXYLIC ACID  
 FT (BY SIMILARITY).  
 FT DISULFID 29 44  
 FT DISULFID 38 50 BY SIMILARITY.  
 FT DISULFID 43 57 BY SIMILARITY.  
 FT DISULFID 61 66 BY SIMILARITY.  
 FT DISULFID 72 87 BY SIMILARITY.  
 FT DISULFID 81 93 BY SIMILARITY.  
 FT DISULFID 86 100 BY SIMILARITY.  
 FT DISULFID 104 109 BY SIMILARITY.  
 FT DISULFID 115 130 BY SIMILARITY.  
 FT DISULFID 124 136 BY SIMILARITY.  
 FT DISULFID 129 143 BY SIMILARITY.  
 FT DISULFID 147 152 BY SIMILARITY.  
 FT DISULFID 158 173 BY SIMILARITY.  
 FT DISULFID 167 179 BY SIMILARITY.  
 FT DISULFID 172 186 BY SIMILARITY.  
 FT DISULFID 190 195 BY SIMILARITY.  
 FT CARBOHYD 206 206 N-LINKED (GLCNAC...)  
 FT SEQUENCE 212 AA: 21209 MW: 80948245DB625A5 CRC64.  
 SQ  
 Query Match 28.1%; Score 52; DB 1; Length 212;  
 Best Local Similarity 45.0%; Pred. No. 5.6;  
 Matches 9; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 15 GVCYDTKITLGNMGYAVPN 34  
 DB 64 GACYTSKRCGTGAGKTCPN 83  
 RESULT 8  
 POLG\_PSBMV STANDARD; PRT; 3206 AA.  
 AC P29152;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Genome polypeptide [contains: N-terminal protein (P1); Helper  
 DE component proteinase (EC 3.4.22.45) (HC-Pro); Protein P3; 6 kDa  
 DE protein 1 (6k1); Cytoplasmic inclusion protein (CI); 6 kDa protein 2

DE (6k2); Genome-linked protein (VPG); Nuclear inclusion protein A (NI-A)  
 DE (NIA) (EC 3.4.22.44) (49 kDa proteinase) (49 kDa-Pro); Nuclear  
 DE inclusion protein B (NI-B) (NIB) (RNA-directed RNA polymerase) (EC  
 DE 2.7.7.48); Coat protein (CP)].  
 OS Pea seed-borne mosaic virus (strain DPDI).  
 OS Viruses: ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
 OC Potyvirus.  
 OX NCBI\_TaxID=31736;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92044431; PubMed=1940858;  
 RA Johansen E., Rasmussen O.F., Heide M., Borkhardt B.;  
 RT "The complete nucleotide sequence of pea seed-borne mosaic virus  
 RNA.";  
 RL J. Gen. Virol. 72:2625-2632(1991).  
 CC -i- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID  
 CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.  
 CC MAY BE INVOLVED IN REPLICATION.  
 CC -i- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT  
 CC -i- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.  
 CC -i- CATALYTIC ACTIVITY: Hydrolyzes glutamyl bonds, and activity is  
 CC further restricted by preferences for the amino acids in p6 - p1,  
 CC that vary with the species of potyvirus, e.g. Glu-Xaa-Xaa-Tyr-Xaa-  
 CC Glu-(Ser or Gly) for the enzyme from tobacco etch virus. The  
 CC natural substrate is the viral polypeptide, but other proteins and  
 CC oligopeptides containing the appropriate consensus sequence are  
 CC also cleaved.  
 CC -i- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(N).  
 CC -i- CATALYTIC ACTIVITY: Hydrolyzes a Gly-I-Gly bond at its own C-  
 CC terminus, commonly in the sequence -Tyr-Xaa-Val-Gly-I-Gly, in the  
 CC processing of the potyviral polypeptide.  
 CC -i- PFM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.  
 CC -i- PFM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE  
 CC POLYPEPTIDE WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC  
 CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT  
 CC INDIVIDUAL PROTEINS.  
 CC -i- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.  
 CC -i- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.  
 CC -i- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPEPTIDE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D10930; BAA01726.1; .  
 CC PIR: J01331; GNVSFV.  
 CC MEROPS: C04.010; .  
 CC MEROPS: C06.001; .  
 CC InterPro: IPR001410; DEAD.  
 CC InterPro: IPR001650; Helicase\_C.  
 CC InterPro: IPR001730; Helicase\_C4.  
 CC InterPro: IPR001456; Peptidase\_C6.  
 CC InterPro: IPR002540; Poly\_P1.  
 CC InterPro: IPR001592; Poly-coat.  
 CC InterPro: IPR001205; RNA\_pol\_P3D.  
 CC InterPro: IPR001254; Ser\_protease\_Try.  
 CC Pfam: PF00271; helicase\_C; 1.  
 CC Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
 CC Pfam: PF00767; Poly-coat; 1.  
 CC Pfam: PF00851; Peptidase\_C6; 1.  
 CC Pfam: PF00863; Peptidase\_C4; 1.  
 CC Pfam: PF01577; Poly\_P1; 1.  
 CC PRINTS: PR00966; NIAPOTYPYPTASE.  
 CC SMART: SM00487; DEXDC; 1.  
 CC SMART: SM00490; HELICC; 1.  
 CC Hydrolase: Transferase; Thiol protease; RNA-directed RNA polymerase;  
 CC Coat protein; Polypeptide; Covalent protein-RNA linkage; Helicase;  
 KW ATP-binding.



DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Leuyl-tRNA synthetase 2 (EC 6.1.1.4) (Leucine--tRNA ligase 2) (LeuRS  
 2)  
 GN LEU2 OR LEU5 OR SSO0589 OR C21-009.  
 OS Sulfolobus solfataricus.  
 OC Archaea: Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 CC Sulfolobus.  
 NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=20165948; PubMed=10701121;  
 RA Charlebois R.L., Singh R.K., Chan-Weher C.C.-Y., Allard G., Chow C.,  
 RA Confalonieri F., Curtis B., Duguet M., Erasuo G., Paquy D.,  
 RA Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,  
 RA Kushiwa N., Lafleur E., Medina N., Peng X., Penny S.L., She Q.,  
 RA St Jean A., van der Oost J., Young F., Zivanovic Y., Doolittle W.F.,  
 RA Ragan M.A., Sensen C.W.;  
 RT "Gene content and organization of a 281-kbp contig from the genome of  
 the extremely thermophilic archaeon, Sulfolobus solfataricus P2.";  
 RL Genome 43:116-136(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Awezaz M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,  
 RA Helamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 RN [3]  
 RP SEQUENCE OF 762-944 FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=97352708; PubMed=9209067;  
 RA Charlebois R.L., Sensen C.W., Doolittle W.F., Brown J.R.;  
 RT "Evolutionary analysis of the hisGABRFDH1 gene cluster from the  
 archaeon Sulfolobus solfataricus P2.";  
 RL J. Bacteriol. 179:4429-4432(1997).  
 CC -I- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +  
 CC diphosphate + L-leucyl-tRNA(Leu).  
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -I- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y18930; CAB5710.1; -  
 DR EMBL: AE006689; AAK40902.1; -  
 DR EMBL: U82227; AAB63016.1; -  
 DR InterPro: IPR004493; Leu-tRNA-synt1a.  
 DR InterPro: IPR002300; tRNA-synt\_1a.  
 DR Pfam: PF00133; tRNA-synt\_1.1.  
 DR TIGRPFAM: TIGR00395; leu5\_arch: 1.  
 DR PROSITE: PS00178; AA-TRNA\_LIGASE\_I, FALSE\_NEG.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Complete proteome.  
 FT SITE 36 46 "HIGH" REGION.  
 FT SITE 623 627 "KMSKS" REGION.  
 FT BINDING 626 626 ATP (BY SIMILARITY).  
 SQ SEQUENCE 944 AA; 109104 MW; 9ACAPCFBF04160DC CRC64;

Query Match

27.6%; Score 51; DB 1; Length 944;

Best Local Similarity 46.2%; Pred. No. 31;  
 Matches 12; Conservative 1; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 GAPVARAEVAPFGVCYFTKLGNN 26  
 DB 301 GTGVMSVPAHAFDFYYSKRTKFN 326  
 RESULT 11  
 ID DHAL\_AGABI STANDARD; PRT; 500 AA.  
 AC 074187;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Aldenhyde dehydrogenase (EC 1.2.1.3) (ALDH).  
 GN ALDA.  
 OS Agaricus bisporus (Common mushroom).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 OC Agaricales; Agaricaceae; Agaricus.  
 NCBI\_TaxID=5341;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Horst H39;  
 RA Schaepe P.J., Muller Y., Visser J.;  
 RT "Molecular structure and spatial expression of housekeeping genes in  
 RT mushrooms";  
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.  
 CC -I- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2O) = an acid + NADH.  
 CC -I- PATHWAY: Ethanol utilization; second step.  
 CC -I- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y17825; CAA76875.1; -  
 DR HSSP: P51977; IBXS.  
 DR InterPro: IPR02086; Aldehyde\_dehydr.  
 DR Pfam: PF00171; aldehyd. 1.  
 DR PROSITE: PS006070; ALDEHYDE\_DEHYDR\_CYS; 1.  
 DR PROSITE: PS00687; ALDEHYDE\_DEHYDR\_GLU; 1.  
 KW Oxidoreductase; NAD.  
 FT NP\_BIND 246 251 NAD (ADP PART) (BY SIMILARITY).  
 FT ACT\_SITE 269 269 BY SIMILARITY.  
 FT ACT\_SITE 303 303 BY SIMILARITY.  
 SQ SEQUENCE 500 AA; 54395 MW; C4FCE58B50855925 CRC64;  
 Query Match 27.3%; Score 50.5; DB 1; Length 500;  
 Best Local Similarity 40.0%; Pred. No. 20;  
 Matches 12; Conservative 4; Mismatches 13; Indels 1; Gaps 1;  
 QY 5 ARRAEVAPEFGVCYFTKLGNNLGYAVPN 34  
 DB 457 ANTIDAGVFPFG-GYKOSGIGRELGAYALNN 485  
 RESULT 12  
 ID GUAA\_AGRT5 STANDARD; PRT; 525 AA.  
 AC 080112;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2) (Glutamine  
 DE amidotransferase) (GMP synthetase).  
 GN GUAA OR ATU0281 OR AGR\_C\_480.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

DT	01-MAY-1992 (Rel. 22, last sequence update)
DT	15-JUN-2002 (Rel. 41, last annotation update)
DE	B-lymphocyte antigen CD19 (B-lymphocyte surface antigen B4) (LEU-12)
DE	(Differentiation antigen CD19) (Fragment).
GN	CD19.
OS	Cavia porcellus (Guinea pig).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Hystricognathi; Cavidae; Cavia.
OX	NCBI_TaxID=10141;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=91332460; PubMed=1714482;
RA	Zhou L.J., Ord D.C., Hughes A.L., Tedder T.F.;
RT	"Structure and domain organization of the CD19 antigen of human,
RT	mouse, and guinea pig B lymphocytes. Conservation of the extensive
RT	cytoplasmic domain."
RU	J. Immunol. 147:1424-1432(1991).
CC	- FUNCTION: MAY BE INVOLVED IN GROWTH REGULATION OF B-CELLS.
CC	- SUBCELLULAR LOCATION: Type I membrane protein.
CC	- SIMILARITY: BELONGS TO THE IMMUGLOBULIN SUPERFAMILY.
CC	- SIMILARITY: CONTAINS 2 IMMUGLOBULIN-LIKE C2-TYPE DOMAINS.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
DR	EMBL: M62543; AAA51529.1; .
DR	InterPro: IPR003599; Ig.
DR	InterPro: IPR003006; Ig_MHC.
DR	Pfam: PF00047; Ig; 1.
DR	SMART: SM00409; IG; 1.
KW	Immunoglobulin domain; Glycoprotein; Transmembrane; B-cell; Repeat.
FT	Non-Ter
FT	DOMAIN 1 168 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 169 190 POTENTIAL.
FT	DOMAIN 191 430 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN <1 17 IG-LIKE C2-TYPE DOMAIN 1.
FT	DOMAIN 69 145 IG-LIKE C2-TYPE DOMAIN 2.
FT	DISELID 76 138 POTENTIAL.
FT	CARBOHYD 1 1 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 14 14 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 28 28 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 59 59 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 142 142 N-LINKED (GLCNAC. . . ) (POTENTIAL).
SQ	SEQUENCE 430 AA: 46957 MW: 84f05B04f94590E3 CRC64:
Query Match	27.0% Score 50; DB 1; Length 430;
Best Local Similarity	27.9%; Pred. No. 20;
Matches 12: Conservative	6; Mismatches 15; Indels 10; Gaps 1;
QY	3 PVARAEEVAPFGVCYDTKTLGNNIGC-----YAEPNV 35     ::   :
Db	347 PVARTTFDLPSPHSCAMPSPREATSLSGSYEDMGILYAAQL 389
RESULT 14	
YK27_CAEEL	
ID YK27_CAEEL STANDARD; PRT; 432 AA.	
AC 016686;	
DT 15-DEC-1998 (Rel. 37, Created)	
DT 15-DEC-1998 (Rel. 37, Last sequence update)	
DT 16-OCT-2001 (Rel. 40, Last annotation update)	
DE Hypothetical 50.0 kDa protein K07E8.7 in chromosome IV.	
GN K07E8.7.	
OS Caenorhabditis elegans.	
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdittida; Rhabdicoidae;	
OC Rhabdittidae; Pelodoriinae; Caenothabditis.	
NN NCBI_TaxID=6239.	
[1]	

RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Jones K., Kramer J.;  
 RL Submitted (Aug-1997) to the EMBL/Genbank/DDI databases.  
 CC -!- SIMILARITY: BELONGS TO THE RLU FAMILY OF PSEUDOURIDINE SYNTHASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF016678; AAB66153.1;  
 CC Wormpep: K07E8.7; CELL1898.  
 CC InterPro: IPR002990; PSL\_RLU.  
 CC InterPro: IPR000613; Pseudou\_synth.  
 CC Pfam: PF00849; Pseudou\_synth\_2; 1.  
 CC TIGRFAMs: TIGR00005; rlud\_subfam; 1.  
 CC PROSITE: PS01129; PSL\_RLU; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 432 AA; 50011 MW; ABC90844BA84734 CRC64;

Query Match 27.0%; Score 50; DB 1; Length 432;  
 Best Local Similarity 36.0%; Pred. No. 20;  
 Matches 9; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

OY 4 VARAEVAPFGVCYDTKTLGNLNG 28  
 DB 60 IVDGVRLAPYMACYRTRTGRWIG 84

RESULT 15  
 SR44 ASPNG  
 ID SR34 ASPNG STANDARD; PRT; 534 AA.  
 AC 000179;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Signal recognition particle 54 kDa protein homolog.  
 GN SRPA.  
 OS Aspergillus niger.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-96144300; PubMed-8566805;  
 RA Thompson S.A., Golightly E.J., Yaver D.S.;  
 RT "Nucleotide sequence of the Aspergillus niger srpa gene."  
 RL Gene 167:337-338(1995).  
 CC -!- FUNCTION: BINDS TO THE SIGNAL SEQUENCE OF PRESECRETORY PROTEIN  
 CC WHEN THEY EMERGE FROM THE RIBOSOMES AND TRANSFERS THEM TO TRAM  
 CC (TRANSLOCATING CHAIN-ASSOCIATING MEMBRANE PROTEIN).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- DOMAIN: HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN BINDS GTP; THE  
 CC M-DOMAIN BINDS THE 7S RNA IN PRESENCE OF SRP19 AND ALSO BINDS THE  
 CC SIGNAL SEQUENCE.  
 CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L38317; AAB04946.1;  
 CC HSSP: 007347; 2FPH.  
 CC InterPro: IPR000897; SRP54.  
 CC InterPro: IPR004125; SRP54\_SPB.

DR Pfam: PF00448; SRP54; 1.  
 DR Pfam: PF02881; SRP54\_N; 1.  
 DR Pfam: PF02978; SRP\_SPB; 1.  
 DR PRODOM: PD000819; SRP54; 1.  
 DR PROSITE: PS00300; SRP54; 1.  
 KW Signal recognition particle: GTP-binding; RNA-binding.  
 FT DOMAIN 1 296  
 FT DOMAIN 297 534  
 FT NP\_BIND 109 116  
 FT NP\_BIND 191 195  
 FT NP\_BIND 249 252  
 FT DOMAIN 468 474  
 FT DOMAIN 493 501  
 FT DOMAIN 522 533  
 FT POLY-GLY.  
 FT POLY-GLY.  
 SQ SEQUENCE 534 AA; 57117 MW; 1B21E3A48CB3BA6 CRC64;

Query Match 27.0%; Score 50; DB 1; Length 534;  
 Best Local Similarity 44.8%; Pred. No. 25;  
 Matches 13; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

OY 7 AVEAVAPFGVCYDTKTLGNLNGYAVPNV 35  
 DB 236 AFKATADFGAIIITTKTDGHAAGGCAISAV 264

Search completed: May 20, 2003, 16:35:35  
 Job time : 12.4231 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:19:09 ; Search time 43.2692 Seconds  
(without alignments)  
166.669 Million cell updates/sec

Title: US-09-869-155-13

Perfect score: 185  
Sequence: 1 GAPVRAVEAVAPGVCYDKTLGNLNGYAVPNV 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeop:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	87	47.0	422	10	08SIV3 oryza sativ
2	79.5	43.0	433	10	09ZVS4 arabidopsis
3	79.5	43.0	434	10	09ZVS5 arabidopsis
4	76	41.1	424	10	08SIV1 oryza sativ
5	74	40.0	419	10	08SIV0 oryza sativ
6	73.5	39.7	452	10	09FSH9 lupinus alb
7	68.5	37.0	413	10	09FSH1 lupinus alb
8	62.5	33.8	449	10	04Z369 lupinus ang
9	59	31.9	433	10	005929 daucus caro
10	58	31.4	402	10	08SIV9 oryza sativ
11	56.5	30.5	525	16	08ZNV0 salmoneilla
12	56.5	30.5	525	16	08ZNV3 salmoneilla
13	56.5	30.5	525	16	08ZNV5 salmoneilla
14	56	30.3	314	10	09FTN5 oryza sativ
15	56	30.3	314	10	08VYH7 oryza sativ
16	55.5	30.0	195	16	09PK21 chlamydia m

17	55.5	30.0	343	7	P79459 ambystoma m
18	55	29.7	369	10	Q9FSZ9 cicer ariet
19	54.5	29.5	594	16	Q9JUB6 neisseria m
20	54	29.2	64	11	Q9JJP3 mus musculu
21	54	29.2	380	10	Q43568 nicotiana t
22	54	29.2	546	10	Q40451 nicotiana t
23	52.5	28.4	238	16	Q8Y2W9 talstonia s
24	52.5	28.4	520	16	Q9ZSQ3 talstonia s
25	52	28.1	242	10	Q9SEZ1 streptomyces
26	52	28.1	340	16	Q9X709 streptomyces
27	52	28.1	433	10	Q8RVH5 glycine max
28	52	28.1	433	10	Q60447 cricetus
29	52	28.1	626	16	Q8XQ86 talstonia s
30	52	28.1	826	10	Q9SCU3 arabidopsis
31	52	28.1	3198	12	Q9INJ4 pea seed-bo
32	51.5	27.8	478	16	Q9RTM9 deinoecoccus
33	51.5	27.8	520	16	Q97PZ3 streptococc
34	51	27.6	310	5	Q76759 anophelis g
35	51	27.6	327	5	Q76758 anophelis g
36	51	27.6	556	4	Q96S68 homo sapien
37	51	27.6	556	4	Q9BRD6 homo sapien
38	51	27.6	800	11	Q63257 rattus norv
39	51	27.6	1048	10	Q8S249 oryza sativ
40	51	27.6	3199	12	Q8S074 pea seed-bo
41	50.5	27.3	336	16	Q8U6P8 agrobacteri
42	50.5	27.3	495	16	Q34660 bacillus su
43	50.5	27.3	525	16	Q8UJ12 agrobacteri
44	50.5	27.3	594	2	Q51188 neisseria m
45	50.5	27.3	867	3	Q96T27 neurospora

## ALIGNMENTS

### RESULT 1

ID 08SIV3 PRELIMINARY: PRT: 422 AA.  
AC 08SIV3;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Putative dermal glycoprotein.  
GN P0504E02.6.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrharioideae; Oryzae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsunoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
clone:P0504E02.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP003269; BAB89705.1.;  
SQ SEQUENCE 422 AA: 44745 MW: 08CF26346400A9F CRC64;

Query Match 47.0%; Score 87; DB 10; Length 422;  
Best local Similarity 56.0%; Pred. No. 0.00097;  
Matches 14; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

OY 11 VAPFGVCYDKTLGNLNGYAVPNV 35  
DB 311 VAPFELCYESRTIGSNRLGAYAPDI 335

### RESULT 2

ID 09ZVS4 PRELIMINARY: PRT: 433 AA.  
AC 09ZVS4;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE F15K9.17 protein (Hypothetical 45.7 kDa protein) (Putative  
 extracellular dermal glycoprotein EDG).  
 GN F15K9.17.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Li J., Liu S.,  
 RA Kremenetskaia I., Luos J., Araujo R., Buehler E., Conway A.B.,  
 RA Dewar K., Feng J., Kim C., Li Y., Shinn P., Sun H., Davis R.W.,  
 RA Ecker J.R., Federpiet N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F15K9 sequence."  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Theologis A.;  
 RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Theologis A.;  
 RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Theologis A.;  
 RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
 RA Palm C.J., Theologis A., Ecker J., Davis R.W.;  
 RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,  
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,  
 RA Yamamuro Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Kim C., Koeseema E., Lam B., Lin J., Meyers M.C., Miranda M.,  
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,  
 RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,  
 RA Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene F15K9.17 (GI:3850579).";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,  
 RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,  
 RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koeseema E.,  
 RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,  
 RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,  
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene F15K9.17 (GI:3850579).";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC005278; AAC72119.1; -;  
 DR EMBL: AF325092; AAK17160.1; -;  
 DR EMBL: AY050988; AAL15204.1; -;  
 DR EMBL: AY035026; AAK59531.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 433 AA: 45717 MW: 7214FC4B8BA72962 CRC64;

Query Match 43.0%; Score 79.5; DB 10; Length 433;  
 Best Local Similarity 47.1%; Pred. No. 0.011;  
 Matches 16; Conservative 6; Mismatches 9; Indels 3; Gaps 1;  
 OY 5 ARAVEAVA--PFGVCYDRTKLTGNLNGYAVPNV 35

DB 315 ARSIRKVASVKEGACFTKRWGVTRLGYAVPEI 348  
 ||::: || |||:: ||::| |||||:  
 RESULT 3  
 ID 09ZVS5 PRELIMINARY; PRT: 434 AA.  
 AC 09ZVS5;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE F15K9.16 protein (Putative extracellular dermal glycoprotein  
 precursor).  
 GN F15K9.16.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Li J., Liu S.,  
 RA Kremenetskaia I., Luos J., Araujo R., Buehler E., Conway A.B.,  
 RA Dewar K., Feng J., Kim C., Li Y., Shinn P., Sun H., Davis R.W.,  
 RA Ecker J.R., Federpiet N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F15K9 sequence."  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Theologis A.;  
 RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Theologis A.;  
 RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Theologis A.;  
 RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Egu P., Lee J.M.,  
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,  
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,  
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene F15K9.16 (GI:3850580).";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC005278; AAC72120.1; -;  
 DR EMBL: AF332411; AAG48774.1; -;  
 SQ SEQUENCE 434 AA: 46148 MW: 17DD7684008FDAFC CRC64;  
 Query Match 43.0%; Score 79.5; DB 10; Length 434;  
 Best Local Similarity 47.1%; Pred. No. 0.011;  
 Matches 16; Conservative 6; Mismatches 9; Indels 3; Gaps 1;  
 OY 5 ARAVEAVA--PFGVCYDRTKLTGNLNGYAVPNV 35  
 ||::: || |||:: ||::| |||||:  
 DB 316 ARSIRKVASVKEGACFTKRWGVTRLGYAVPEI 349  
 ||::: || |||:: ||::| |||||:  
 RESULT 4  
 ID 08S1V1 PRELIMINARY; PRT: 424 AA.  
 AC 08S1V1;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Putative dermal glycoprotein.  
 GN P0504E02.8.

```
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare(GAS) genomic DNA, chromosome 1, PAC
RT clone: P0504E02."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003269; BAB89707.1;
SQ SEQUENCE 424 AA; 44620 MW; 0DACD5C8C852C34 CRC64;

Query Match
Best Local Similarity 54.1%; Score 76; DB 10; Length 424;
Matches 13; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 12 AFGVCYDRTKLTGNLGGYAVPNV 35
DB 313 APPELCYDSRDLGSTRIGYAVPQI 336

RESULT 5
Q8S1V0 PRELIMINARY; PRT; 419 AA.
AC Q8S1V0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative dermal glycoprotein.
GN P0504E02.9.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare(GAS) genomic DNA, chromosome 1, PAC
RT clone: P0504E02."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003269; BAB89708.1;
SQ SEQUENCE 419 AA; 43786 MW; 792BE55A4F3F3A8 CRC64;

Query Match
Best Local Similarity 40.0%; Score 74; DB 10; Length 419;
Matches 16; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 8 VEVAPEGVCYDRTKLTGNLGGYAVPNV 35
DB 307 VAAVAFPLCYRASMLGNTRIGYEPVAV 334

RESULT 6
Q9FSH9 PRELIMINARY; PRT; 452 AA.
AC Q9FSH9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Conglutin gamma precursor.
OS Lupinus albus (White lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OX NCBI_TaxID=3870;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ULTRA;
```

```
RA Scarafoni A., Di Cataldo A., Vassilevskaia T.D., Bekman E.P.,
RA Rodrigues-Pousada C., Cecilian F., Duranti M.;
RT "Cloning, sequencing and expression in seeds and radicles of two
RT Lupinus albus conglutin gamma genes."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ297490; CAC16394.1;
DR InterPro; IPR001461; AsproteaseA1.
DR Pfam; PF00026; asp; 1.
KW Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 452 CONGLUTIN GAMMA.
SQ SEQUENCE 452 AA; 49219 MW; C343BA283BE752FF CRC64;

Query Match
Best Local Similarity 39.7%; Score 73.5; DB 10; Length 452;
Matches 17; Conservative 4; Mismatches 7; Indels 5; Gaps 2;

QY 3 PVARAVEAVAFGVCYDRTKLTGNLGGYAVPNV 35
DB 339 PKQAVKAVGPFGLCYDKTKIS--GG--VPSV 366

RESULT 7
Q9FEX1 PRELIMINARY; PRT; 413 AA.
AC Q9FEX1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Conglutin gamma (Fragment).
OS Lupinus albus (White lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OX NCBI_TaxID=3870;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ULTRA;
RA Scarafoni A., Di Cataldo A., Vassilevskaia T.D., Bekman E.P.,
RA Rodrigues-Pousada C., Cecilian F., Duranti M.;
RT "Cloning, sequencing and expression in seeds and radicles of two
RT Lupinus albus conglutin gamma genes."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ297568; CAC17729.1;
FT NON_TER 1 1
SQ SEQUENCE 413 AA; 45732 MW; 022723A38B16E4B7 CRC64;

Query Match
Best Local Similarity 37.0%; Score 68.5; DB 10; Length 413;
Matches 15; Conservative 6; Mismatches 7; Indels 5; Gaps 2;

QY 3 PVARAVEAVAFGVCYDRTKLTGNLGGYAVPNV 35
DB 300 PKDAOVESVGFGLCYDSRKL--GG--IPSV 327

RESULT 8
Q42369 PRELIMINARY; PRT; 449 AA.
AC Q42369;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Conglutin gamma precursor.
OS Lupinus angustifolius (Narrow-leaved blue lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OX NCBI_TaxID=3871;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VAR. UNICROP; TISSUE=COFFLEDON;
RA Gayler K.R.;
```

```
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-VAR. UNICROP; TISSUE=COTYLEDON;
RX MEDLINE=93144715; PubMed=8425065;
RA Koliyas S., Gayler K.R.;
RT "Structure of the cdna coding for conglutin gamma, a sulphur-rich
RL plant mol. Biol. 21:397-401(1993).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN-VAR. UNICROP; TISSUE=COTYLEDON;
RX Gayler K.R.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RX MEDLINE=95388773; PubMed=7659758;
FA Millar A.A., Kunst L.;
RL Plant Physiol. 108:1747-1747(1995).
CR EMBL; X65601; CAA46552.1; -.
DR EMBL; L39786; AAB53771.1; -.
DR InterPro: IPR001461; AsproteaseA1.
DR Pfam: PF00026; asp_1.
KA Signal.
FT SIGNAL.
FT CHAIN 33 449 POTENTIAL.
SQ SEQUENCE 449 AA; 48916 MW; 348BE0CFBD084F3 CRC64;
CONGLUTIN GAMMA.

Query Match
Best Local Similarity 42.4%; Score 62.5; DB 10; Length 449;
Matches 14; Conservative 6; Mismatches 8; Indels 5; Gaps 2;

OY 3 PVAFAEAVAFPGVCYDTKTLGNNGYAVPNV 35
DB 336 PKOAVKAVGPFGLCYDSRKIS---GC-APSV 363

RESULT 9
O05929 PRELIMINARY; PRT; 433 AA.
ID 005929;
AC 005929;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE EDGP precursor (Fragment).
GN EDGPI.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX NCBL_TaxID=4039;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=ROOT;
RA Saton S.;
RL Submitted (MAR-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE=ROOT;
RA Satch S., Sturm A., Fujii T., Chrispeels M.J.;
RT "cdna cloning of an extracellular dermal glycoprotein of carrot and
RL its expression in response to wounding.";
RL Planta 188:432-438(1992).
DR EMBL; D14550; BAA03413.1; -.
KW Signal.
FT NON_TER 1 1
FT SIGNAL 20 POTENTIAL.
FT CHAIN 21 433 EDGP.
SQ SEQUENCE 433 AA; 45841 MW; 06EDFC19B7BDC38E CRC64;

Query Match
Best Local Similarity 39.3%; Score 59; DB 10; Length 433;
Matches 11; Conservative 7; Mismatches 10; Indels 0; Gaps 0;
```

```
OY 8 VEAAPFGVCYDTKTLGNNGYAVPNV 35
DB 323 VASVAPFGACFTDNILSTRLGPSVPSI 350

RESULT 10
O08109 PRELIMINARY; PRT; 402 AA.
ID 08109;
AC 08109;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Putative dermal glycoprotein.
GN P0504E02.10.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzaceae; Oryza.
OX NCBL_TaxID=39947;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RL clone: P0504E02."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003269; BAB89709.1; -.
SQ SEQUENCE 402 AA; 41793 MW; E8C84A214D6F6720 CRC64;

Query Match
Best Local Similarity 47.8%; Score 58; DB 10; Length 402;
Matches 11; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 10 AVAPFGVCYDTKTLGNNGYAV 32
DB 294 SVAPFELCENSSALGFTRVGYAV 316

RESULT 11
O08260 PRELIMINARY; PRT; 525 AA.
ID 08260;
AC 08260;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE GMP synthetase (EC 6.3.5.2).
GN GUA OR STM2510.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBL_TaxID=602;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGGC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of salmonella enterica serovar Typhimurium
RL LT2."
RL Nature 413:852-856(2001).
DR EMBL; AE008813; ARL21404.1; -.
DR InterPro: IPR000991; GATase_1.
DR InterPro: IPR004739; GMPsynthase_N.
DR InterPro: IPR001674; GMP_synth_C.
DR InterPro: IPR004135; tRNA_Me_trans.
DR Pfam: PF00117; GATase; 1.
DR Pfam: PF00958; GMP_synth_C; 1.
DR Pfam: PF03054; tRNA_Me_trans; 1.
DR PRINTS; PR00096; GATASE.
```

DR TIGRFAMS: TIGR00884; guaa\_Cterm; 1.  
 DR TIGRFAMS: TIGR00888; guaa\_Nterm; 1.  
 DR PROSITE: PS00442; GATASE\_TYPE\_I; 1.  
 KW Ligase: Complete proteome.  
 SQ SEQUENCE 525 AA; 58716 MW; D354406650969744 CRC64;

Query Match 30.5%; Score 56.5; DB 16; Length 525;  
 Best Local Similarity 52.2%; Pred. No. 20;  
 Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

OY 9 EAVAP-FGVCYDRTKLTGNNGY 30  
 DB 77 EAGVPVFGVCYGMOTMAMQLGSH 99

## RESULT 12

O82403 PRELIMINARY; PRT; 525 AA.  
 AC 082403;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE GMP synthase (glutamine-hydrolyzing) (EC 6.3.5.2).  
 GN STY2751.  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=601.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed-11677608;  
 RA Parkhill J., Dougan G., James K.D., Houlden N.R., Pickard D., Wain J.,  
 Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 Krog A., Larsen T.S., Leather A., Mole S., O'Gaora P., Parry C.,  
 Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 Whitehead S., Barrett B.G.;  
 RA "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).  
 DR EMBL: AL627275; CAD02712.1; -  
 DR InterPro: IPR000991; GATase\_1.  
 DR InterPro: IPR004739; GMPsynthase\_N.  
 DR InterPro: IPR001674; GMP\_synth\_C.  
 DR InterPro: IPR004135; tRNA\_Me\_trans.  
 DR Pfam: PF00117; GATase; 1.  
 DR Pfam: PF00958; GMP\_synth\_C; 1.  
 DR Pfam: PF03054; tRNA\_Me\_trans; 1.  
 DR PRINTS: PR00096; GATASE.  
 DR TIGRFAMS: TIGR00884; guaa\_Cterm; 1.  
 DR TIGRFAMS: TIGR00888; guaa\_Nterm; 1.  
 DR PROSITE: PS00442; GATASE\_TYPE\_I; 1.  
 KW Ligase: Complete proteome.  
 SQ SEQUENCE 525 AA; 58686 MW; 3086B517F41EA9F9 CRC64;

Query Match 30.5%; Score 56.5; DB 16; Length 525;  
 Best Local Similarity 52.2%; Pred. No. 20;  
 Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

OY 9 EAVAP-FGVCYDRTKLTGNNGY 30  
 DB 77 EAGVPVFGVCYGMOTMAMQLGSH 99

## RESULT 13

O8XAB5 PRELIMINARY; PRT; 525 AA.  
 AC 08XAB5;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE GMP synthase (glutamine-hydrolyzing).  
 GN GUAA OR 23771 OR ECS3369.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed-11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 Grobeck E.J., Davis N.W., Lim A., Dialanta E.T., Potomousis K.,  
 Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed-11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 Kuhara S., Shiba T., Hattori M., Shingawa H.;  
 RA "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL: AE005480; AAG57618.1; -  
 DR EMBL: AP002561; BAB36792.1; -  
 DR InterPro: IPR000991; GATase\_1.  
 DR InterPro: IPR004739; GMPsynthase\_N.  
 DR InterPro: IPR001674; GMP\_synth\_C.  
 DR InterPro: IPR004135; tRNA\_Me\_trans.  
 DR Pfam: PF00117; GATase; 1.  
 DR Pfam: PF00958; GMP\_synth\_C; 1.  
 DR Pfam: PF03054; tRNA\_Me\_trans; 1.  
 DR PRINTS: PR00096; GATASE.  
 DR TIGRFAMS: TIGR00884; guaa\_Cterm; 1.  
 DR TIGRFAMS: TIGR00888; guaa\_Nterm; 1.  
 DR PROSITE: PS00442; GATASE\_TYPE\_I; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 525 AA; 58665 MW; 33077AA286E60171 CRC64;

Query Match 30.5%; Score 56.5; DB 16; Length 525;  
 Best Local Similarity 52.2%; Pred. No. 20;  
 Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

OY 9 EAVAP-FGVCYDRTKLTGNNGY 30  
 DB 77 EAGVPVFGVCYGMOTMAMQLGSH 99

## RESULT 14

O9FTN5 PRELIMINARY; PRT; 314 AA.  
 AC O9FTN5;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE putative isoflavone reductase homolog IRL.  
 GN P0005A05.14.  
 OS Oryza sativa (Rice).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eubacteriidae; Oryzaeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone: p0005a05.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP002863; BAB16910.1; -;  
 DR InterPro: IPR000875; Cecropin.  
 DR InterPro: IPR000595; CNMP\_binding.  
 DR InterPro: IPR003866; Isoflav\_reduct.  
 DR Pfam: PF02716; Isoflavone\_redu; 1.  
 DR PROSITE: PS00268; CECROPIN; UNKNOWN.1.  
 DR PROSITE: PS00888; CNMP\_BINDING\_1; UNKNOWN.1.  
 SO SEQUENCE 314 AA; 33501 MW; C6E33803F2E2018 CRC64;

Query Match 30.3%; Score 56; DB 10; Length 314;  
 Best Local Similarity 47.1%; Pred. No. 14;  
 Matches 16; Conservative 3; Mismatches 11; Indels 4; Gaps 2;

OY 2 APVARRAEVAPFGVCYDTKLGNNLGGYAVPNV 35  
 DB 141 AAVRAVEAA--GVPY-TYVSNYFAGYALPTI 170

## RESULT 15

C8VYH7 PRELIMINARY; PRT; 314 AA.  
 ID Q8VYH7;  
 AC Q8VYH7;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 D3 Isoflavone reductase-like protein.  
 O3 Oryza sativa (rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=4530;  
 RX [1]  
 RF SEQUENCE FROM N.A.  
 RA Kim S.T., Kim S.G., Yu S., Young Y.C., Kang K.Y.;  
 RT "Isolation and Characterization of a Rice Isoflavone Reductase (OSIRF)  
 RT Gene Expressed by Rice Blast Fungal Elicitor.";  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY071920; AAL61542.1; -;  
 DR InterPro: IPR000875; Cecropin.  
 DR InterPro: IPR000595; CNMP\_binding.  
 DR InterPro: IPR003866; Isoflav\_reduct.  
 DR Pfam: PF02716; Isoflavone\_redu; 1.  
 DR PROSITE: PS00268; CECROPIN; UNKNOWN.1.  
 DR PROSITE: PS00888; CNMP\_BINDING\_1; UNKNOWN.1.  
 SO SEQUENCE 314 AA; 33487 MW; 534D958039ZCB6C1 CRC64;

Query Match 30.3%; Score 56; DB 10; Length 314;  
 Best Local Similarity 47.1%; Pred. No. 14;  
 Matches 16; Conservative 3; Mismatches 11; Indels 4; Gaps 2;

OY 2 APVARRAEVAPFGVCYDTKLGNNLGGYAVPNV 35  
 DB 141 AAVRAVEAA--GVPY-TYVSNYFAGYALPTI 170

Search completed: May 20, 2003, 16:43:07  
 Job time : 46.2692 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:35:02 ; Search time 16.1538 seconds  
(without alignments)  
63.750 Million cell updates/sec

Title: US-09-869-155-13

Perfect score: 185  
Sequence: 1 GAPVARAVEAVAPFCVCDTKTLGNLNGYAVPNV 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	29.2	546	2	US-08-669-721-3
2	54	29.2	546	4	US-09-189-344-3
3	51.5	27.8	516	4	US-09-197-063-2
4	50	27.0	534	2	US-08-317-401E-2
5	50	27.0	552	2	US-08-317-401E-4
6	49	26.5	635	4	US-09-341-833A-6
7	49	26.5	635	4	US-09-341-833A-7
8	49	26.5	635	4	US-09-341-833A-9
9	49	26.5	677	4	US-09-341-833A-8
10	46.5	25.1	783	1	US-08-101-593-6
11	46.5	25.1	783	1	US-08-465-995A-6
12	46.5	25.1	783	2	US-08-465-995A-6
13	46	24.9	332	4	US-09-105-390-54
14	46	24.9	1091	3	US-08-633-768A-2
15	46	24.9	1940	4	US-08-644-271-30
16	46	24.9	1940	4	US-09-077-955-34
17	45.5	24.6	393	4	US-09-629-616-2
18	45.5	24.6	523	4	US-09-550-338-2
19	45	24.3	626	4	US-08-961-083-220
20	45	24.3	1121	1	US-07-789-915A-2
21	45	24.3	1121	1	US-08-005-002C-2
22	45	24.3	1121	1	US-08-487-203A-2
23	45	24.3	1334	6	US-07-908-253-3
24	45	24.3	1403	2	US-08-694-865-17
25	45	24.3	1403	2	US-08-535-837-3
26	45	24.3	1403	2	US-08-535-837-3
27	45	24.3	1403	3	US-09-124-491-17

28	44	23.8	332	4	US-09-105-390-58	Sequence 58, Appl
29	44	23.8	399	3	US-08-651-999A-3	Sequence 3, Appl
30	44	23.8	399	4	US-09-385-752-3	Sequence 3, Appl
31	44	23.8	731	2	US-08-911-364-1	Sequence 1, Appl
32	44	23.8	733	4	US-08-464-700-2	Sequence 2, Appl
33	44	23.8	765	4	US-09-134-001C-3681	Sequence 3681, Ap
34	44	23.8	792	2	US-08-678-039A-40	Sequence 40, Appl
35	44	23.8	15281	2	US-08-471-119A-2	Sequence 2, Appl
36	43.5	23.5	332	4	US-09-064-411A-39	Sequence 39, Appl
37	43.5	23.5	366	4	US-09-064-411A-37	Sequence 37, Appl
38	43	23.2	212	2	US-08-663-566A-7	Sequence 7, Appl
39	43	23.2	212	2	US-08-023-610-7	Sequence 7, Appl
40	43	23.2	212	2	US-08-288-065A-7	Sequence 7, Appl
41	43	23.2	212	2	US-08-362-240A-7	Sequence 7, Appl
42	43	23.2	212	4	US-08-804-372A-5	Sequence 5, Appl
43	43	23.2	212	5	PCT-US95-10345-7	Sequence 7, Appl
44	43	23.2	331	4	US-09-105-390-56	Sequence 56, Appl
45	43	23.2	370	1	US-07-966-187-12	Sequence 12, Appl

#### ALIGNMENTS

```
RESULT 1
US-08-669-721-3
; Sequence 3, Application US/08669721
; Patent No. 5834236
;
GENERAL INFORMATION:
; APPLICANT: Lamb et al., Christopher J.
; TITLE OF INVENTION: NOVEL TRANSCRIPTION ENHANCER ELEMENT AND
; TITLE OF INVENTION: TRANSCRIPTION FACTOR AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
;
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,721
; FILING DATE: 27-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ellison, Eldora L.
; REGISTRATION NUMBER: 39,967
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-669-721-3
Query Match 29.2% Score 54; DB 2; Length 546;
Best Local Similarity 35.3%; Pred. No. 5.8;
Matches 12; Conservative 3; Mismatches 19; Indels 0; Gaps 0;
OY 1 GAPVARAVEAVAPFCVCDTKTLGNLNGYAVPN 34
DB 400 GPPATVGVTDVPIAARDTEMLPNAVGCGVTN 433
RESULT 2
```

US-09-189-344-3  
; Sequence 3, Application US/09189344  
; Patent No. 6191258  
; GENERAL INFORMATION:  
; APPLICANT: Lamb et al., Christopher J.  
; TITLE OF INVENTION: NOVEL TRANSCRIPTION ENHANCER ELEMENT AND  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR AND METHODS OF USE THEREFOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/189,344  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/669,721  
; FILING DATE: 27-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ellison, Eldora L.  
; REGISTRATION NUMBER: 39,967  
; REFERENCE/DOCKET NUMBER: 07251/014001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 546 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-189-344-3

Query Match 29.2%; Score 54; DB 4; Length 546;  
Best Local Similarity 35.3%; Pred. No. 5.8;  
Matches 12; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

QY 1 GAVPAAVEAVAPFGVCYDTKLTGNNLGGYAVPN 34  
DB 400 GPPATVGVTDVPIAAPTETMLPNAVGGGVYN 433

RESULT 3  
US-09-197-063-2  
; Sequence 2, Application US/09197063  
; Patent No. 6261817  
; GENERAL INFORMATION:  
; APPLICANT: Zalacain, Magdalena  
; APPLICANT: Brown, James R.  
; APPLICANT: Biswas, Sanjoy  
; APPLICANT: Warren, Richard L.  
; APPLICANT: Shilling, Lisa K.  
; TITLE OF INVENTION: No. 6261817el Guaa  
; FILE REFERENCE: GM10121  
; CURRENT APPLICATION NUMBER: US/09/197,063  
; CURRENT FILING DATE: 1998-11-20  
; EARLIER APPLICATION NUMBER: 60/066,350  
; NUMBER OF SEQ ID NOS: 5  
; SCFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 516  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae

US-09-197-063-2  
Query Match 27.8%; Score 51.5; DB 4; Length 516;  
Best Local Similarity 33.3%; Pred. No. 13;  
Matches 11; Conservative 4; Mismatches 9; Indels 9; Gaps 1;

QY 1 GAVPAAVEAVAPFGVCYDTKLTGNNLGGYAVPN 33  
DB 79 GPI-----LGICGMQLTTHKLGKVP 102

RESULT 4  
US-08-317-401E-2  
; Sequence 2, Application US/08317401E  
; Patent No. 5922561  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Sheryl Ann  
; APPLICANT: Yaver, Debbie Sue  
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF  
; TITLE OF INVENTION: ASPERGILLUS NIGER  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5922561o No. 5922561disk of No. 5922561th America, Inc.  
; STREET: 405 Lexington Avenue, Suite 6400  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,401E  
; FILING DATE: 03-October-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Harrington, James J.  
; REGISTRATION NUMBER: 38,711  
; REFERENCE/DOCKET NUMBER: 4248.000-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 867 0123  
; TELEFAX: 212 867 0298  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 534 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Aspergillus niger  
; US-08-317-401E-2

Query Match 27.0%; Score 50; DB 2; Length 534;  
Best Local Similarity 44.8%; Pred. No. 23;  
Matches 13; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 7 AVEAVAPFGVCYDTKLTGNNLGGYAVPNV 35  
DB 236 ARKATADFGAIIITTKDGHAGGAI SAV 264

RESULT 5  
US-08-317-401E-4  
; Sequence 4, Application US/08317401E  
; Patent No. 5922561  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Sheryl Ann  
; APPLICANT: Yaver, Debbie Sue  
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF  
; TITLE OF INVENTION: ASPERGILLUS NIGER  
; NUMBER OF SEQUENCES: 10





Sequence 8, Application US/09341833A  
Patent No. 6383488  
GENERAL INFORMATION:  
APPLICANT: CENTRO DE INGENIERIA GENETICA Y BIOTECNOLOGIA  
TITLE OF INVENTION: PRE-M/M PROTEIN EPITOPES OF THE DENGUE VIRUS, SYNTHETIC  
FILE REFERENCE: Dengue prem/M SEOS 1-9 for 976-4(0003)  
CURRENT APPLICATION NUMBER: US/09/341,833A  
PRIOR FILING DATE: 1999-10-14  
PRIOR APPLICATION NUMBER: CU 13/97  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 677  
TYPE: PRF  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Fusion Protein  
OTHER INFORMATION: of Dengue virus prem/M and Neisseria meningitidis  
US-09-341-833A-8

Query Match 26.5%; Score 49; DB 4; Length 677;  
Best Local Similarity 37.5%; Pred. No. 43;  
Matches 15; Conservative 2; Mismatches 13; Indels 10; Gaps 2;

QY 2 GAPVAREAVAP-----FG-----VCYDFTKLTGNNLGGYVA 31  
DB 175 GAPVAREAVAPAAAPAAQFGSGADAEYDVVVLGGPGGYS 214

RESULT 10  
US-08-101-593-6  
Sequence 6, Application US/08101593  
Patent No. 5547859  
GENERAL INFORMATION:  
APPLICANT: Goodman, Myron F.  
APPLICANT: Reha-Krantz, Linda J.  
TITLE OF INVENTION: NEW DNA SEQUENCING ENZYMES  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Robbins, Berliner & Carson  
STREET: 201 No. 5547859th Figueroa Street, Fifth Floor  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90012-2628  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/101,593  
FILING DATE: 19930802  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Spitals, John P.  
REGISTRATION NUMBER: 29,215  
REFERENCE/DOCKET NUMBER: 1920-305  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 977-1001  
TELEFAX: (213) 977-1003  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 783 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-101-593-6

Query Match 25.1%; Score 46.5; DB 1; Length 783;

Best Local Similarity 37.1%; Pred. No. 1.2e+02;  
Matches 13; Conservative 4; Mismatches 15; Indels 3; Gaps 1;  
QY 1 GAPVAREAVAPFGCYDFTKLTGNNLGGYVAPNV 35  
DB 360 GLPVDHGGSVAAFGHLTFPRM--HRAGYVAPNL 391

RESULT 11  
US-08-465-995A-6  
Sequence 6, Application US/08465995A  
Patent No. 5660980  
GENERAL INFORMATION:  
APPLICANT: Myron F. Goodman  
APPLICANT: Linda J. Reha-Krantz  
TITLE OF INVENTION: METHODS FOR IDENTIFYING AND ISOLATING  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Robbins, Berliner & Carson, LLP  
STREET: 201 No. 5660980th Figueroa Street, Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90012-2628  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: ASCII DOS/TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,995A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Robert Berliner  
REGISTRATION NUMBER: 20,121  
REFERENCE/DOCKET NUMBER: 1920-30502  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 213/977-1001  
TELEFAX: 213/977-1003  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 783 base pairs  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-465-995A-6

Query Match 25.1%; Score 46.5; DB 1; Length 783;  
Best Local Similarity 37.1%; Pred. No. 1.2e+02;  
Matches 13; Conservative 4; Mismatches 15; Indels 3; Gaps 1;

QY 1 GAPVAREAVAPFGCYDFTKLTGNNLGGYVAPNV 35  
DB 360 GLPVDHGGSVAAFGHLTFPRM--HRAGYVAPNL 391

RESULT 12  
US-08-465-994C-6  
Sequence 6, Application US/08465994C  
Patent No. 5928919  
GENERAL INFORMATION:  
APPLICANT: MYRON F. GOODMAN  
APPLICANT: LINDA J. REHA-KRANTZ  
TITLE OF INVENTION: VARIANT DNA POLYMERASES  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Robbins, Berliner & Carson, LLP  
STREET: 201 No. 5928919th Figueroa Street, Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90012-2628



```

RESULT 15
US-08-644-271-30
: Sequence 30 Application US/08644271
: Patent No. 5814478
:
: GENERAL INFORMATION:
: APPLICANT: Valenzuela, et al.
: TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
: TITLE OF INVENTION: AND LIGANDS
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Regeneron Pharmaceuticals, Inc.
: STREET: 777 Old Saw Mill Road
: CITY: Tarrytown
: STATE: NY
: COUNTRY: USA
: ZIP: 10591
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ version 2.0
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/644,271
: FILING DATE: 10-MAY-1996
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: USSN 60/008,657
: FILING DATE: 15-DEC-1995
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Covert, Robert J
: REGISTRATION NUMBER: 36,108
: REFERENCE/DOCKET NUMBER: REG 195A
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 914-345-7400
: TELEFAX: 914-345-7721
:
: TELEX:
:
: INFORMATION FOR SEQ ID NO: 30:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1940 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FEATURE:
: NAME/KEY: Rat Agrin
: LOCATION: 1..1940
: OTHER INFORMATION:
:
: US-08-644-271-30

```

Search completed: May 20, 2003, 16:44:37  
Job time : 18.1538 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:49:49 ; Search time 17.3077 Seconds  
(without alignments)  
200.549 Million cell updates/sec

Title: US-09-869-155-13  
Perfect score: 185  
Sequence: 1 GAPVAAVEAAVAPFCVCYDRTKLNLCGAVAPNV 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCF\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCFUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56.5	30.5	525	10 US-09-815-242-10249	Sequence 10249, A
2	56.5	30.5	525	10 US-09-815-242-11741	Sequence 11741, A
3	56.5	30.5	525	10 US-09-815-242-13950	Sequence 13950, A
4	54	29.2	546	10 US-09-839-743-3	Sequence 3, Appli
5	52.5	28.4	525	10 US-09-815-242-11952	Sequence 11952, A
6	52	28.1	142	9 US-09-738-626-6400	Sequence 6400, Ap
7	51.5	27.8	520	10 US-09-815-242-13283	Sequence 13283, A
8	49.5	26.8	520	10 US-09-815-242-10605	Sequence 10605, A
9	48	25.9	267	9 US-09-813-408-22	Sequence 22, Appl
10	48	25.9	712	10 US-09-815-242-10930	Sequence 10930, A
11	47.5	25.7	523	10 US-09-815-242-10987	Sequence 10987, A
12	46.5	25.1	719	9 US-10-270-875-34	Sequence 34, Appl
13	46.5	25.1	719	9 US-10-270-878-34	Sequence 34, Appl
14	46.5	25.1	719	9 US-10-270-786-34	Sequence 34, Appl
15	46.5	25.1	719	9 US-10-270-710-34	Sequence 34, Appl
16	46.5	25.1	719	9 US-10-270-859-34	Sequence 34, Appl
17	46.5	25.1	1092	9 US-09-423-126-5	Sequence 5, Appli
18	24.9	26.6	10	US-09-837-235-18	Sequence 18, Appl
19	46	24.9	588	9 US-09-908-193-34	Sequence 34, Appl

20	46	24.9	695	10 US-09-764-898-191	Sequence 191, App
21	46	24.9	778	9 US-09-908-193-8	Sequence 8, Appli
22	46	24.9	779	9 US-09-908-193-10	Sequence 10, Appli
23	46	24.9	1091	9 US-09-423-126-2	Sequence 2, Appli
24	46	24.9	1091	10 US-09-280-197-2	Sequence 2, Appli
25	46	24.9	1940	9 US-10-016-283-34	Sequence 34, Appl
26	45.5	24.6	177	9 US-09-791-279-148	Sequence 148, App
27	45.5	24.6	393	9 US-09-738-626-5280	Sequence 5280, Ap
28	45.5	24.6	393	9 US-10-284-138-2	Sequence 2, Appli
29	45.5	24.6	393	9 US-10-284-334-2	Sequence 2, Appli
30	45.5	24.6	393	10 US-09-836-470B-2	Sequence 2, Appli
31	45.5	24.6	523	9 US-09-738-626-4177	Sequence 4177, Ap
32	45.5	24.6	523	10 US-09-767-878-2	Sequence 2, Appli
33	45	24.3	577	10 US-09-758-269-10	Sequence 10, Appl
34	45	24.3	626	10 US-09-765-272-220	Sequence 220, App
35	45	24.3	1010	10 US-09-759-152-2	Sequence 2, Appli
36	45	24.3	1024	10 US-09-879-257A-30	Sequence 30, Appl
37	45	24.3	5179	9 US-10-025-380-1068	Sequence 1068, Ap
38	45	24.3	5179	10 US-09-922-217-1068	Sequence 1068, Ap
39	45	24.3	5179	10 US-09-833-263-1068	Sequence 1068, Ap
40	44.5	24.1	103	9 US-09-796-692-830	Sequence 830, App
41	44.5	24.1	103	9 US-09-796-692-2420	Sequence 2420, Ap
42	44.5	24.1	103	9 US-10-040-862-830	Sequence 830, App
43	44.5	24.1	103	9 US-10-040-862-2420	Sequence 2420, Ap
44	44.5	24.1	203	10 US-09-809-545A-2	Sequence 2, Appli
45	44.5	24.1	372	10 US-09-873-880-40	Sequence 40, Appli

## ALIGNMENTS

RESULT 1  
US-09-815-242-10249  
Sequence 10249, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA-011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FASTED for Windows Version 4.0  
SEQ ID NO 10249  
LENGTH: 525  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-815-242-10249  
Query Match 30.5%; Score 56.5; DB 10; Length 525;  
Best Local Similarity 52.2%; Pred. No. 3.4;

Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Qy 9 EAVAP-FGVCYDTKTLGNLCGY 30  
||| ||||| :|: |||:  
Db 77 EAGVPEFGVCYGMOTMAMOLGCH 99

RESULT 2  
US-09-815-242-11741  
Sequence 11741, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11741  
LENGTH: 525  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)...(525)  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-815-242-11741

Query Match 30.5%; Score 56.5; DB 10; Length 525;  
Best Local Similarity 52.2%; Pred. No. 3.4;  
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Qy 9 EAVAP-FGVCYDTKTLGNLCGY 30  
||| ||||| :|: |||:  
Db 77 EAGVPEFGVCYGMOTMAMOLGCH 99

RESULT 3  
US-09-815-242-13950  
Sequence 13950, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13950  
LENGTH: 525  
TYPE: PRT  
ORGANISM: Salmonella typhi  
US-09-815-242-13950

Query Match 30.5%; Score 56.5; DB 10; Length 525;  
Best Local Similarity 52.2%; Pred. No. 3.4;  
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Qy 9 EAVAP-FGVCYDTKTLGNLCGY 30  
||| ||||| :|: |||:  
Db 77 EAGVPEFGVCYGMOTMAMOLGCH 99

RESULT 4  
US-09-839-743-3  
Sequence 3, Application US/09839743  
Patent No. US20020146824A1  
GENERAL INFORMATION:  
APPLICANT: The Salk Institute for Biological Sciences  
APPLICANT: Lamb, Christopher  
APPLICANT: Doerner, Peter  
APPLICANT: Labile, Goetz  
TITLE OF INVENTION: NO. US20020146824A1el Transcription Enhancer Element and  
FILE REFERENCE: SALKINS.008DV3  
CURRENT APPLICATION NUMBER: US/09/839,743  
PRIOR FILING DATE: 2001-04-19  
PRIOR APPLICATION NUMBER: US 09/401,336  
PRIOR FILING DATE: 1999-09-21  
PRIOR APPLICATION NUMBER: US 09/189,344  
PRIOR FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: US 08/669,721  
PRIOR FILING DATE: 1996-06-27  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 546  
TYPE: PRT  
ORGANISM: Nicotiana tabacum  
US-09-839-743-3

Query Match 29.2%; Score 54; DB 10; Length 546;  
Best Local Similarity 35.3%; Pred. No. 8.5;  
Matches 12; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

Qy 1 GAVVAVAVAVAPFGVCYDTKTLGNLCGYAVPN 34  
||| ||||| :|: |||:  
Db 400 GPPAATVGVTVPIAAAFDTENLNNVAGGGCVTN 433



```

: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011a
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10605
: LENGTH: 520
: TYPE: PRF
: ORGANISM: Enterococcus faecalis
US-09-815-242-10605

Query Match          26.8%  Score 49.5; DB 10; Length 520;
Best Local Similarity 30.3%  Pred. No. 38;
Matches 10; Conservative 5; Mismatches 9; Indels 9; Gaps 1;

OY 1 GAFVARAEVAPFGVCYDTKTLGNNGYAVP 33
Db 82 GIPD-----LGICYGMQMLTNHNGKVEP 105

RESULT 9
US-09-813-408-22
: Sequence 22, Application US/09813408
: Publication No. US20030049619A1
: GENERAL INFORMATION:
: APPLICANT: Delagrave, Simon
: APPLICANT: Mafts, Barry
: TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial L
: FILE REFERENCE: HER0041
: CURRENT APPLICATION NUMBER: US/09/813,408
: CURRENT FILING DATE: 2001-03-21
: NUMBER OF SEQ ID NOS: 85
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 22
: LENGTH: 267
: TYPE: PRF
: ORGANISM: Bacillus amyloliquefaciens
US-09-813-408-22

Query Match          25.9%  Score 48; DB 9; Length 267;
Best Local Similarity 44.8%  Pred. No. 29;
Matches 13; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

OY 1 GAFVARAEVAPFGVCYDTKTLGNNG 29
Db 66 GTHVAGTVAAVASASLAVKVLGADGSG 94

RESULT 10
```

```

US-09-815-242-10930
: Sequence 10930, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011a
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10930
: LENGTH: 712
: TYPE: PRF
: ORGANISM: Enterococcus faecalis
US-09-815-242-10930

Query Match          25.9%  Score 48; DB 10; Length 712;
Best Local Similarity 43.5%  Pred. No. 90;
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

OY 1 GAFVARAEVAPFGVCYDTKTL 23
Db 170 GTLYAKAEVKEVPEEIAFDRTL 192

RESULT 11
US-09-815-242-10987
: Sequence 10987, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011a
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
```



```

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10987
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-10987

Query Match
Best Local Similarity 34.5%; Score 47.5; DB 10; Length 523;
Matches 10; Conservative 3; Mismatches 7; Indels 9; Gaps 1;

QY 1 GAPVAAVEAVAPFGVCYDTKTLGNLGG 29
    ||| 1:1 1:1 |||
Db 78 GVPV-----LGTCYGMQTMAMQLGG 97

RESULT 12
US-10-270-875-34
; Sequence 34, Application US/10270875
; Publication No. US20030082741A1
; GENERAL INFORMATION:
; APPLICANT: Sigridur Hjorleifsdottir
; APPLICANT: Gudmundur O. Hreggvidsson
; APPLICANT: Olafur H. Fridjonsson
; APPLICANT: Arnthor Aevarsson
; APPLICANT: Jakob K. Kristjansson
; TITLE OF INVENTION: Bacteriophage RM378 of a Thermophilic
; FILE REFERENCE: 2739.1001-001
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/585,858
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/137,120
; PRIOR FILING DATE: 1999-06-02
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-270-875-34

Query Match
Best Local Similarity 37.1%; Score 46.5; DB 9; Length 719;
Matches 13; Conservative 4; Mismatches 15; Indels 3; Gaps 1;

QY 1 GAPVAAVEAVAPFGVCYDTKTLGNLGGYAVPNV 35
    ||| 1:1 1:1 |||
Db 308 GLPVDHRGGVAAFGHLTFPRM---HRAGYVAPNL 339

RESULT 13
US-10-270-878-34
; Sequence 34, Application US/10270878
; Publication No. US20030082790A1
; GENERAL INFORMATION:
; APPLICANT: Sigridur Hjorleifsdottir
; APPLICANT: Gudmundur O. Hreggvidsson
; APPLICANT: Olafur H. Fridjonsson
; APPLICANT: Arnthor Aevarsson
; APPLICANT: Jakob K. Kristjansson
; TITLE OF INVENTION: Bacteriophage RM378 of a Thermophilic
; FILE REFERENCE: 2739.1001-001
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/585,858
```

```

; CURRENT APPLICATION NUMBER: US/10/270,878
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/585,858
; PRIOR FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-270-878-34

Query Match
Best Local Similarity 37.1%; Score 46.5; DB 9; Length 719;
Matches 13; Conservative 4; Mismatches 15; Indels 3; Gaps 1;

QY 1 GAPVAAVEAVAPFGVCYDTKTLGNLGGYAVPNV 35
    ||| 1:1 1:1 |||
Db 308 GLPVDHRGGVAAFGHLTFPRM---HRAGYVAPNL 339

RESULT 14
US-10-270-786-34
; Sequence 34, Application US/10270786
; Publication No. US20030087392A1
; GENERAL INFORMATION:
; APPLICANT: Sigridur Hjorleifsdottir
; APPLICANT: Gudmundur O. Hreggvidsson
; APPLICANT: Olafur H. Fridjonsson
; APPLICANT: Arnthor Aevarsson
; APPLICANT: Jakob K. Kristjansson
; TITLE OF INVENTION: Bacteriophage RM378 of a Thermophilic
; FILE REFERENCE: 2739.1001-001
; CURRENT APPLICATION NUMBER: US/10/270,786
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/585,858
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/137,120
; PRIOR FILING DATE: 1999-06-02
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-270-786-34

Query Match
Best Local Similarity 37.1%; Score 46.5; DB 9; Length 719;
Matches 13; Conservative 4; Mismatches 15; Indels 3; Gaps 1;

QY 1 GAPVAAVEAVAPFGVCYDTKTLGNLGGYAVPNV 35
    ||| 1:1 1:1 |||
Db 308 GLPVDHRGGVAAFGHLTFPRM---HRAGYVAPNL 339

RESULT 15
US-10-270-710-34
; Sequence 34, Application US/10270710
; Publication No. US20030092128A1
; GENERAL INFORMATION:
; APPLICANT: Sigridur Hjorleifsdottir
; APPLICANT: Gudmundur O. Hreggvidsson
; APPLICANT: Olafur H. Fridjonsson
; APPLICANT: Arnthor Aevarsson
; APPLICANT: Jakob K. Kristjansson
; TITLE OF INVENTION: Bacteriophage RM378 of a Thermophilic
; FILE REFERENCE: 2739.1001-001
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/585,858
```

; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: 60/137,120  
; PRIOR FILING DATE: 1999-06-02  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO: 34  
; LENGTH: 719  
; TYPE: PRF  
; ORGANISM: Escherichia coli  
US-10-270-710-34

Query Match 25.1%; Score 46.5; DB 9; Length 719;  
Best Local Similarity 37.1%; Pred. NO. 1.5e+02;  
Matches 13; Conservative 4; Mismatches 15; Indels 3; Gaps 1;

QY 1 GAPVAVAEVAPFGVCYCDTKLTGNNLGGYAVPNV 35  
DB 308 GLPVDRHGGSVAAPFGHLTFPRM--HRAGIVAPNL 339

Search completed: May 20, 2003, 18:00:00  
Job time: 18.3077 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:04:12 ; Search time 18.4945 Seconds  
(without alignments)  
122.483 Million cell updates/sec

Title: US-09-869-155-14

Perfect score: 96

Sequence: 1 KRLGFSRLPHTGCGGL 17

Scoring table:

BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	17	21	AAV93757
2	96	100.0	17	22	AAE14688
3	96	100.0	17	22	AAU07393
4	96	100.0	196	23	AAU75817
5	96	100.0	370	23	AAU75815
6	96	100.0	370	23	AAU75816
7	96	100.0	381	23	AAU75812
8	96	100.0	381	23	AAU75813
9	96	100.0	381	23	AAU75814
10	61	63.5	14	23	AAU75829

11	46.5	48.4	198	21	AAG24723
12	46.5	48.4	198	21	AAG45304
13	46.5	48.4	252	21	AAG24722
14	46.5	48.4	252	21	AAG45303
15	46.5	48.4	256	21	AAG24721
16	46.5	48.4	256	21	AAG45302
17	45	46.9	233	22	AAE14745
18	45	46.9	276	23	AAE14746
19	45	46.9	280	23	AAE14743
20	45	46.9	289	23	AAE14744
21	45	46.9	395	23	AAE14744
22	45	46.9	397	23	AAE17308
23	45	46.9	399	22	AAE17308
24	45	46.9	399	23	AAU77496
25	45	46.9	473	22	AAU40252
26	45	46.9	656	22	ABG25839
27	45	46.9	666	21	ABN42538
28	45	46.9	731	23	AAE14746
29	44.5	46.4	218	22	AAE14746
30	44.5	46.4	218	22	AAE14746
31	44	45.8	477	22	ABG12357
32	43.5	45.3	158	21	AAV58264
33	43.5	45.3	438	21	AAV58262
34	43.5	45.3	472	21	AAV58266
35	43	44.8	163	23	ABG04853
36	42	43.8	159	22	AAU04665
37	42	43.8	163	23	ABG04848
38	42	43.8	325	21	AAE23196
39	42	43.8	325	21	AAE32391
40	42	43.8	463	21	AAE07155
41	42	43.8	494	21	AAE48101
42	42	43.8	844	22	ABG08453
43	42	43.8	841	21	AAE65359
44	41	42.7	58	22	AAU54989
45	41	42.7	58	22	AAU54989

#### ALIGNMENTS

RESULT 1	
AAV93757	
ID	AAV93757 standard; peptide: 17 AA.
AC	AAV93757;
DT	03-OCT-2000 (first entry)
DE	Amino acid sequence of an endo-beta-1,4-xylanase inhibitor peptide.
XX	
XX	
KW	Endo-beta-1,4-xylanase inhibitor; wheat flour; xylanase; bakery product;
KW	dough; dough preparation.
XX	
OS	Triticum sp.
XX	
PN	WO200039289-A2.
XX	
PD	06-JUL-2000.
XX	
XX	
PF	17-DEC-1999; 99WO-1B02071.
XX	
PR	23-DEC-1998; 98GB-0028599.
PR	06-APR-1999; 98GB-0007805.
PR	15-APR-1999; 99GB-0008645.
XX	
PA	(DANI-) DANISCO AS.
PI	Sibbesen O, Sorensen JF;
XX	
DR	WPI; 2000-465744/40.
XX	
PT	Mutant xylanase protein identified using xylanase inhibitor useful for preparing non-sticky dough for bakery products -

XX Claim 24; Page 112; 112pp; English.  
 PS  
 CC The present sequence is derived from an endo-beta-1,4-xylanase  
 CC inhibitor. The protein is obtained from wheat flour. The specification  
 CC also describes a mutant xylanase protein. The xylanase is useful for  
 CC preparing a foodstuff, preferably a bakery product or a substance  
 CC (e.g. a dough) for making the bakery product. Wild type xylanase or  
 CC mutant xylanase is useful for preparing a dough that is less sticky  
 CC than a dough comprising a fungal xylanase. The xylanase inhibitor is  
 CC useful for screening high degree resistance xylanases for dough  
 CC preparation. The xylanase is also useful for preparing a non-sticky  
 CC dough. A combination of xylanase and the inhibitor is useful for  
 CC calibrating and/or determining the quantity of inhibitor in a wheat  
 CC flour sample.  
 CC  
 SO Sequence 17 AA;  
 XX  
 XX  
 Query Match 100.0%; Score 96; DB 21; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KRIGFSRLPHFTGCGGL 17  
 DB 1 KRIGFSRLPHFTGCGGL 17  
 RESULT 2  
 AAE14688  
 ID AAE14688 standard; peptide; 17 AA.  
 AC AAE14688;  
 XX  
 DT 21-AUG-2002 (first entry)  
 XX  
 DE wheat flour xylanase inhibitor A chain C-terminal fragment.  
 XX  
 KW refrigerated dough; syruping; arabinoxylan; bakery product; bread;  
 KW pizza base; cake; biscuit; wheat; flour; xylanase inhibitor.  
 OS Trifolium aestivum.  
 XX  
 PN WO200152657-A1.  
 PD 26-JUL-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-1B00168.  
 XX  
 PR 18-JAN-2000; 2000GB-0001136.  
 XX  
 PA (DANI-) DANISCO AS.  
 XX  
 PI Poulsen CH, Sorensen JF;  
 XX  
 DR WPI; 2001-457446/49.  
 XX  
 PT Production of refrigerated dough with reduced syruping, useful in  
 PT production of bakery products such as bread, comprises admixing cereal  
 PT flour, water and protein that prevents enzymatic degradation of  
 PT arabinoxylan in the cereal flour -  
 XX  
 PS Disclosure; Page 23; 26pp; English.  
 XX  
 CC The invention relates to a process for producing refrigerated dough  
 CC with reduced 'syruping' (precipitation of liquid on the dough surface  
 CC because of a reduction in water holding capacity caused by the breakdown  
 CC of arabinoxylan over time). The process comprises admixing cereal flour  
 CC and water with a protein that reduces/prevents enzymatic degradation of  
 CC arabinoxylan in the cereal flour. The preferred protein is a xylanase  
 CC inhibitor. The method is useful to produce refrigerated dough in which  
 CC syruping is reduced or eliminated. Refrigerated dough is typically  
 CC stored for long periods to enable fresh baked products (e.g. bread,  
 CC pizza bases, cakes, biscuits) to be produced as required. It may also be

CC possible to produce dough or baked products tailored to specific  
 CC requirements using the method by the use of specific proteins/protein  
 CC combinations. The present sequence is wheat flour  
 CC endo-beta-1,4-xyranase inhibitor A chain C-terminal fragment.  
 XX  
 SO Sequence 17 AA;  
 XX  
 XX  
 \* Query Match 100.0%; Score 96; DB 22; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KRIGFSRLPHFTGCGGL 17  
 DB 1 KRIGFSRLPHFTGCGGL 17  
 RESULT 3  
 AA007393  
 ID AA007393 standard; protein; 17 AA.  
 XX  
 AC AA007393;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Bacillus subtilis xylanase inhibitor #2.  
 XX  
 KW xylanase; plant cell wall; baking; cereal; starch production; wood;  
 KW wood pulp bleaching; hemicellulose; arabinoxylan; foodstuff; inhibitor.  
 XX  
 OS Bacillus subtilis.  
 XX  
 PN WO200166711-A1.  
 PD 13-SEP-2001.  
 XX  
 PF 08-MAR-2001; 2001WO-1B00426.  
 XX  
 PR 08-MAR-2000; 2000GB-0005585.  
 PR 27-JUN-2000; 2000GB-0015751.  
 XX  
 PA (DANI-) DANISCO AS.  
 XX  
 PI Sibbesen O, Sorensen JF;  
 XX  
 DR WPI; 2001-596834/67.  
 XX  
 PT Novel variant xylanase polypeptide or its fragment useful for degrading  
 PT or modifying plant cell wall, comprises amino acid modifications such  
 PT that the polypeptide has altered sensitivity to xylanase inhibitor -  
 XX  
 PS Disclosure; Page 63; 70pp; English.  
 XX  
 CC The invention relates to a variant xylanase polypeptide (I) or its  
 CC fragment having xylanase activity, comprising one or more amino acid  
 CC modifications such that (I) or its fragment has an altered sensitivity to  
 CC a xylanase inhibitor as compared with the parent xylanase enzyme. (I) or  
 CC its coding sequence (II) is useful for degrading or modifying plant cell  
 CC wall or for processing a plant material by contacting the plant cell wall  
 CC or plant material with (I) or (II). (I) is useful for modifying plant  
 CC materials, and in baking, processing cereals, starch production,  
 CC processing wood and enhancing the bleaching of wood pulp. (II)  
 CC is useful for altering the viscosity derived from the presence of  
 CC hemicellulose or arabinoxylan in a solution or system comprising plant  
 CC cell wall material. (I) is useful for preparing a foodstuff such as  
 CC bread, pretzels, tortillas, cakes, cookies, biscuits or crackers.  
 CC The present sequence represents the amino acid sequence of Bacillus  
 CC subtilis xylanase inhibitor #2 as described in the method of the  
 CC invention.  
 XX  
 SO Sequence 17 AA;  
 XX  
 XX  
 Query Match 100.0%; Score 96; DB 22; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-08;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KRIGSRPLPHFGCGGL 17  
 1 KRIGSRPLPHFGCGGL 17

Db 1 KRIGSRPLPHFGCGGL 17

RESULT 4  
 AAU75817  
 ID AAU75817 standard: Protein; 196 AA.  
 AC AAU75817;  
 XX  
 XX 23-APR-2002 (first entry)  
 XX

Wheat L endoxylanase inhibitor. TDXI I, partial sequence TDXI-I.01.

Wheat; TDXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 immobilised enzyme; enzyme; dough syruuping; cereal product; beer; plant;  
 straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 noodle; animal feed; starch separation; maize processing; malting;  
 plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

Triticum durum cultivar Mexicali.

WO200198474-A1.

27-DEC-2001.

21-JUN-2001; 2001WO-BE00106.

22-JUN-2000; 2000GB-0015296.  
 25-JAN-2001; 2001GB-0002018.  
 26-JAN-2001; 2001GB-0002194.  
 16-MAR-2001; 2001GB-0006564.  
 21-MAY-2001; 2001GB-0012328.

PR  
 PA (LEUV-) LEUVEN RES & DEV.  
 XX  
 XX Delcour J, Debysier W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR N-PSDB: ABK13674.  
 XX  
 XX WPI: 2002-114579/15.

Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 beta-glucanolytic enzymes comprising using endoxylanases during  
 screening for inhibition activity or affinity chromatography with  
 immobilised enzymes

Claim 127; Page 57; 127pp; English.

The invention relates to separating and/or isolating inhibitors of  
 cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 screening the inhibition activity by using two or more enzymes during the  
 separation and/or isolation steps that allow to distinguish inhibitors of  
 different specificity or by using an affinity chromatographic step with  
 immobilised enzymes and/or antibodies against inhibitors. Also  
 included are an isolated nucleic acid molecule encoding an inhibitor  
 which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 nucleic acid, an expression system transformed with the nucleic  
 acid, a host organism transformed with the nucleic acid, the inhibitory  
 proteins encoded by the nucleic acids and modulators of the proteins.  
 A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 plant or plant materials transformed with the nucleic acid are useful  
 for the formation of an endoxylanase-inhibitor complex, screening  
 endoxylanases that are totally, less or not inhibited by the inhibitors,  
 reducing syruuping in refrigerated dough compositions, affecting the  
 relative affinity and/or relative hydrolysis specificity and/or relative  
 hydrolysis rate versus water-extractable and/or water-unextractable

arabinoxylans of endoxylanases such as by the formation of an  
 endoxylanase/inhibitor complex, improving the malting of cereals such as  
 barley, sorghum and wheat and/or the production of beer, improving the  
 production and/or quality of baked or extruded cereal products such as  
 straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 biscuits, pasta and noodles, animal feed stuff, improving the production  
 of starch derived syrups, sorbitol, xylitol, xylose and/or xylitol, wheat gluten  
 starch separation and production, improving maize processing,  
 plant disease resistance and nutraceutical and/or pharmaceutical  
 applications, improving paper and pulp technologies. The present  
 sequence represents a partial wheat TDXI-I (T. durum L endoxylanase  
 inhibitor).

Sequence 196 AA:

Query Match 100.0%; Score 96; DB 23; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KRIGSRPLPHFGCGGL 17  
 180 KRIGSRPLPHFGCGGL 196

Db 180 KRIGSRPLPHFGCGGL 196

RESULT 5  
 AAU75815  
 ID AAU75815 standard: Protein; 370 AA.  
 AC AAU75815;  
 XX  
 XX 23-APR-2002 (first entry)  
 XX

Wheat L endoxylanase inhibitor. TAXI I, partial sequence TAXI-I.0.

Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 immobilised enzyme; enzyme; dough syruuping; cereal product; beer; plant;  
 straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 noodle; animal feed; starch separation; maize processing; malting;  
 plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

Triticum aestivum cultivar Solissons.

WO200198474-A1.

27-DEC-2001.

21-JUN-2001; 2001WO-BE00106.

22-JUN-2000; 2000GB-0015296.  
 25-JAN-2001; 2001GB-0002018.  
 26-JAN-2001; 2001GB-0002194.  
 16-MAR-2001; 2001GB-0006564.  
 21-MAY-2001; 2001GB-0012328.

PR  
 PA (LEUV-) LEUVEN RES & DEV.  
 XX  
 XX Delcour J, Debysier W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR N-PSDB: ABK13672.  
 XX  
 XX WPI: 2002-114579/15.

Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 beta-glucanolytic enzymes comprising using endoxylanases during  
 screening for inhibition activity or affinity chromatography with  
 immobilised enzymes

Claim 127; Page 57; 127pp; English.

The invention relates to separating and/or isolating inhibitors of  
 cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 screening the inhibition activity by using two or more enzymes during the

CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitor  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syrruping in rehydrated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial wheat TAXI-I (T. aestivum L endoxylanase  
 CC inhibitor).

SO Sequence 370 AA:

Query Match 100.0%; Score 96; DB 23; Length 370;

Best Local Similarity 100.0%; Pred. No. 3.5e-07;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KRIGFSRLPHFTGCCGL 17

DB 354 KRIGFSRLPHFTGCCGL 370

RESULT 6  
 AAU75816  
 ID AAU75816 standard; Protein: 370 AA.  
 AC AAU75816;  
 XX  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE wheat L endoxylanase inhibitor, TAXI I, partial sequence TAXI-I.02.  
 XX  
 XX wheat; TAXI-I; L endoxylanase inhibitor; cellulytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syrruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Trifolium aestivum cultivar Estica.  
 XX  
 XX  
 PN WO200198474-A1.  
 PD 27-DEC-2001.  
 XX  
 XX 21-JUN-2001; 2001WO-BE00106.  
 PF  
 XX  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 23-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.

XX Delcour J, Debysse W, Gebruers K, Goesaert H, Flerens K, Robben J;  
 PI Van Campenhout S;  
 XX  
 DR WPI: 2002-114579/15.  
 DR N-PSDB: ABK13673.  
 XX  
 XX Separating and/or isolating inhibitors of cellulytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes -  
 XX  
 PS Claim 127; Page 57; 127pp; English.

XX The invention relates to separating and/or isolating inhibitors of  
 CC cellulytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitor  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syrruping in rehydrated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial wheat TAXI-I (T. aestivum L endoxylanase  
 CC inhibitor).

SO Sequence 370 AA:

Query Match 100.0%; Score 96; DB 23; Length 370;

Best Local Similarity 100.0%; Pred. No. 3.5e-07;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KRIGFSRLPHFTGCCGL 17

DB 354 KRIGFSRLPHFTGCCGL 370

RESULT 7  
 AAU75812  
 ID AAU75812 standard; Protein: 381 AA.  
 AC AAU75812;  
 XX  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 XX wheat L endoxylanase inhibitor, TAXI I, variant #1.  
 DE  
 XX  
 XX wheat; TAXI-I; L endoxylanase inhibitor; cellulytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syrruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW

KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX Trifolium aestivum.  
 OS  
 XX  
 FT Key Location/Qualifiers  
 FT Misc-difference 95  
 FT Misc-difference /label= Unknown  
 FT Misc-difference 98  
 FT Misc-difference /label= Unknown  
 FT Misc-difference 101  
 FT Misc-difference /label= Unknown  
 FT Misc-difference 110  
 FT Misc-difference /label= Unknown  
 FT Misc-difference 333  
 FT Misc-difference /label= Unknown  
 XX  
 PN WO200198474-A1.  
 XX  
 XX 27-DEC-2001.  
 PD  
 XX 21-JUN-2001; 2001WO-BE00106.  
 PF  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 XX (LEUV-) LEUVEN RES & DEV.  
 PA  
 XX Delcour J, Debysers W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 PI  
 XX MPI; 2002-114579/15.  
 DR  
 XX Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 50-51; 127pp; English.  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex. screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syripping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex. improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, wheat xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a variant of wheat TAXI-I (T. aestivum L. endoxylanase

CC inhibitor).  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.  
 XX  
 SQ Sequence 381 AA;  
 Query Match 100.0%; Score 96; DB 23; Length 381;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-07;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KRLGESRLPHFTGCGGL 17  
 Db 365 KRLGESRLPHFTGCGGL 381  
 RESULT 8  
 AA075813  
 ID AA075813 standard; Protein; 381 AA.  
 XX  
 AC AA075813;  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Wheat L endoxylanase inhibitor, TAXI I, variant #2.  
 XX  
 XX Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syripping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Trifolium aestivum.  
 XX  
 FH Key Location/Qualifiers  
 FH Misc-difference 95  
 FT Misc-difference /label= Unknown  
 FT Misc-difference 98  
 FT Misc-difference /label= Unknown  
 FT Misc-difference 101  
 FT Misc-difference /label= Unknown  
 FT Misc-difference 110  
 FT Misc-difference /label= Unknown  
 FT Misc-difference 333  
 FT Misc-difference /label= Unknown  
 XX  
 PN WO200198474-A1.  
 XX  
 XX 27-DEC-2001.  
 PD  
 XX 21-JUN-2001; 2001WO-BE00106.  
 PF  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 XX (LEUV-) LEUVEN RES & DEV.  
 PA  
 XX Delcour J, Debysers W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 PI  
 XX MPI; 2002-114579/15.  
 DR  
 XX Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 51; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of







PR 30-APR-1999: 99US-0132048.  
PR 30-APR-1999: 99US-0132407.  
PR 04-MAY-1999: 99US-0132484.  
PR 05-MAY-1999: 99US-0132485.  
PR 06-MAY-1999: 99US-0132486.  
PR 06-MAY-1999: 99US-0132487.  
PR 07-MAY-1999: 99US-0132487.  
PR 11-MAY-1999: 99US-0132486.  
PR 14-MAY-1999: 99US-0134218.  
PR 14-MAY-1999: 99US-0134219.  
PR 14-MAY-1999: 99US-0134221.  
PR 14-MAY-1999: 99US-0134370.  
PR 18-MAY-1999: 99US-0134768.  
PR 19-MAY-1999: 99US-0134941.  
PR 20-MAY-1999: 99US-0135124.  
PR 21-MAY-1999: 99US-0135353.  
PR 24-MAY-1999: 99US-0135629.  
PR 25-MAY-1999: 99US-0136021.  
PR 27-MAY-1999: 99US-0136392.  
PR 28-MAY-1999: 99US-0136782.  
PR 01-JUN-1999: 99US-0137222.  
PR 03-JUN-1999: 99US-0137528.  
PR 04-JUN-1999: 99US-0137502.  
PR 07-JUN-1999: 99US-0137724.  
PR 08-JUN-1999: 99US-0138094.  
PR 10-JUN-1999: 99US-0138540.  
PR 10-JUN-1999: 99US-0138847.  
PR 14-JUN-1999: 99US-0139119.  
PR 16-JUN-1999: 99US-0139452.  
PR 16-JUN-1999: 99US-0139453.  
PR 16-JUN-1999: 99US-0139454.  
PR 17-JUN-1999: 99US-0139492.  
PR 18-JUN-1999: 99US-0139454.  
PR 18-JUN-1999: 99US-0139455.  
PR 18-JUN-1999: 99US-0139456.  
PR 18-JUN-1999: 99US-0139457.  
PR 18-JUN-1999: 99US-0139458.  
PR 18-JUN-1999: 99US-0139459.  
PR 18-JUN-1999: 99US-0139460.  
PR 18-JUN-1999: 99US-0139461.  
PR 18-JUN-1999: 99US-0139462.  
PR 18-JUN-1999: 99US-0139463.  
PR 18-JUN-1999: 99US-0139750.  
PR 18-JUN-1999: 99US-0139763.  
PR 21-JUN-1999: 99US-0139817.  
PR 22-JUN-1999: 99US-0139899.  
PR 23-JUN-1999: 99US-0140353.  
PR 23-JUN-1999: 99US-0140354.  
PR 24-JUN-1999: 99US-0140695.  
PR 28-JUN-1999: 99US-0140823.  
PR 29-JUN-1999: 99US-0140991.  
PR 30-JUN-1999: 99US-0141287.  
PR 01-JUL-1999: 99US-0141842.  
PR 01-JUL-1999: 99US-0142154.  
PR 02-JUL-1999: 99US-0142055.  
PR 06-JUL-1999: 99US-0142390.  
PR 08-JUL-1999: 99US-0142803.  
PR 09-JUL-1999: 99US-0142920.  
PR 12-JUL-1999: 99US-0142977.  
PR 13-JUL-1999: 99US-0143542.  
PR 14-JUL-1999: 99US-0143624.  
PR 15-JUL-1999: 99US-0144005.  
PR 16-JUL-1999: 99US-0144085.  
PR 16-JUL-1999: 99US-0144086.  
PR 19-JUL-1999: 99US-0144325.  
PR 19-JUL-1999: 99US-0144331.  
PR 19-JUL-1999: 99US-0144332.  
PR 19-JUL-1999: 99US-0144333.  
PR 19-JUL-1999: 99US-0144334.  
PR 19-JUL-1999: 99US-0144335.  
PR 20-JUL-1999: 99US-0144352.  
PR 20-JUL-1999: 99US-0144352.  
PR 20-JUL-1999: 99US-0144684.  
PR 21-JUL-1999: 99US-0144814.  
  
PR 21-JUL-1999: 99US-0145086.  
PR 21-JUL-1999: 99US-0145088.  
PR 22-JUL-1999: 99US-0145085.  
PR 22-JUL-1999: 99US-0145087.  
PR 22-JUL-1999: 99US-0145089.  
PR 22-JUL-1999: 99US-0145089.  
PR 23-JUL-1999: 99US-0145192.  
PR 23-JUL-1999: 99US-0145145.  
PR 23-JUL-1999: 99US-0145218.  
PR 23-JUL-1999: 99US-0145224.  
PR 26-JUL-1999: 99US-0145276.  
PR 27-JUL-1999: 99US-0145813.  
PR 27-JUL-1999: 99US-0145918.  
PR 27-JUL-1999: 99US-0145919.  
PR 28-JUL-1999: 99US-0145951.  
PR 02-AUG-1999: 99US-0146386.  
PR 02-AUG-1999: 99US-0146388.  
PR 02-AUG-1999: 99US-0146388.  
PR 02-AUG-1999: 99US-0146389.  
PR 03-AUG-1999: 99US-0147038.  
PR 04-AUG-1999: 99US-0147204.  
PR 04-AUG-1999: 99US-0147302.  
PR 05-AUG-1999: 99US-0147192.  
PR 05-AUG-1999: 99US-0147260.  
PR 06-AUG-1999: 99US-0147303.  
PR 06-AUG-1999: 99US-0147303.  
PR 09-AUG-1999: 99US-0147416.  
PR 09-AUG-1999: 99US-0147493.  
PR 10-AUG-1999: 99US-0147935.  
PR 10-AUG-1999: 99US-0148171.  
PR 11-AUG-1999: 99US-0148319.  
PR 12-AUG-1999: 99US-0148341.  
PR 13-AUG-1999: 99US-0148341.  
PR 13-AUG-1999: 99US-0148341.  
PR 13-AUG-1999: 99US-0148341.  
PR 16-AUG-1999: 99US-0148684.  
PR 17-AUG-1999: 99US-0149368.  
PR 18-AUG-1999: 99US-0149175.  
PR 18-AUG-1999: 99US-0149426.  
PR 20-AUG-1999: 99US-0149722.  
PR 20-AUG-1999: 99US-0149723.  
PR 20-AUG-1999: 99US-0149729.  
PR 23-AUG-1999: 99US-0149902.  
PR 23-AUG-1999: 99US-0149930.  
PR 25-AUG-1999: 99US-0150566.  
PR 26-AUG-1999: 99US-0150884.  
PR 27-AUG-1999: 99US-0151065.  
PR 27-AUG-1999: 99US-0151065.  
PR 27-AUG-1999: 99US-0151066.  
PR 27-AUG-1999: 99US-0151080.  
PR 30-AUG-1999: 99US-0151303.  
PR 31-AUG-1999: 99US-0151438.  
PR 01-SEP-1999: 99US-0151930.  
PR 07-SEP-1999: 99US-0152363.  
PR 10-SEP-1999: 99US-0153070.  
PR 13-SEP-1999: 99US-0153758.  
PR 15-SEP-1999: 99US-0154018.  
PR 16-SEP-1999: 99US-0154039.  
PR 20-SEP-1999: 99US-0154779.  
PR 22-SEP-1999: 99US-0155139.  
PR 23-SEP-1999: 99US-0155486.  
PR 24-SEP-1999: 99US-0155659.  
PR 28-SEP-1999: 99US-0156458.  
PR 29-SEP-1999: 99US-0156596.  
PR 04-OCT-1999: 99US-0157117.  
PR 05-OCT-1999: 99US-0157153.  
PR 06-OCT-1999: 99US-0157865.  
PR 07-OCT-1999: 99US-0158029.  
PR 08-OCT-1999: 99US-0158232.  
PR 12-OCT-1999: 99US-0158369.  
PR 13-OCT-1999: 99US-0159293.  
PR 13-OCT-1999: 99US-0159294.  
PR 13-OCT-1999: 99US-0159295.  
PR 14-OCT-1999: 99US-0159329.  
PR 14-OCT-1999: 99US-0159330.  
PR 14-OCT-1999: 99US-0159331.  
PR 14-OCT-1999: 99US-0159637.  
PR 14-OCT-1999: 99US-0159638.  
PR 18-OCT-1999: 99US-0159584.



PR	23-JUL-1999	9905-01455214
PR	23-JUL-1999	9905-01455224
PR	26-JUL-1999	9905-01455276
PR	27-JUL-1999	9905-01455913
PR	27-JUL-1999	9905-01455918
PR	27-JUL-1999	9905-01455919
PR	28-JUL-1999	9905-01455919
PR	02-AUG-1999	9905-01465386
PR	02-AUG-1999	9905-01465388
PR	02-AUG-1999	9905-01465688
PR	06-AUG-1999	9905-01473003
PR	06-AUG-1999	9905-01473165
PR	09-AUG-1999	9905-01474733
PR	09-AUG-1999	9905-01474793
PR	10-AUG-1999	9905-01481711
PR	11-AUG-1999	9905-01483319
PR	12-AUG-1999	9905-01483431
PR	13-AUG-1999	9905-01485855
PR	13-AUG-1999	9905-01486864
PR	16-AUG-1999	9905-01493368
PR	17-AUG-1999	9905-01494175
PR	18-AUG-1999	9905-01494126
PR	20-AUG-1999	9905-01497122
PR	20-AUG-1999	9905-01497123
PR	20-AUG-1999	9905-01499029
PR	23-AUG-1999	9905-01499320
PR	23-AUG-1999	9905-01499320
PR	25-AUG-1999	9905-01505066
PR	26-AUG-1999	9905-01506884
PR	27-AUG-1999	9905-01510665
PR	27-AUG-1999	9905-01510665
PR	27-AUG-1999	9905-01510880
PR	30-AUG-1999	9905-01513438
PR	31-AUG-1999	9905-01513438
PR	01-SEP-1999	9905-01513303
PR	07-SEP-1999	9905-01526363
PR	22-SEP-1999	9905-01554136
PR	23-SEP-1999	9905-01555189
PR	24-SEP-1999	9905-01556539
PR	28-SEP-1999	9905-01564588
PR	29-SEP-1999	9905-01565696
PR	04-OCT-1999	9905-01571117
PR	05-OCT-1999	9905-01577133
PR	06-OCT-1999	9905-01577535
PR	07-OCT-1999	9905-01578655
PR	08-OCT-1999	9905-01580232
PR	12-OCT-1999	9905-01582632
PR	13-OCT-1999	9905-01583839
PR	13-OCT-1999	9905-01589364
PR	13-OCT-1999	9905-01592934
PR	13-OCT-1999	9905-01592925
PR	14-OCT-1999	9905-01593229
PR	14-OCT-1999	9905-01593330
PR	14-OCT-1999	9905-01593330
PR	18-OCT-1999	9905-01596338
PR	18-OCT-1999	9905-01595884
PR	21-OCT-1999	9905-01607041
PR	21-OCT-1999	9905-01607041
PR	21-OCT-1999	9905-01607668
PR	21-OCT-1999	9905-01607768
PR	21-OCT-1999	9905-01608114
PR	21-OCT-1999	9905-01608115
PR	22-OCT-1999	9905-01609800

[illegible]

PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142390.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154479.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157573.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 990S-0161361.  
PR 28-OCT-1999; 990S-0161920.  
PR 28-OCT-1999; 990S-0161922.  
PR 28-OCT-1999; 990S-0161933.  
PR 29-OCT-1999; 990S-0162142.  
Query Match 48.4%; Score 46.5; DB 21; Length 252;  
Best Local Similarity 43.5%; Pred. No. 18;  
Matches 10; Conservative 1; Mismatches 3; Indels 9; Gaps 1;  
QY 2 RLGSRP-----HFTGCG 15  
Db 61 RVSFRRPGSVSLQRRHFLGCG 83  
RESULT 14  
ID AAG45303 standard; Protein; 252 AA.  
AC AAG45303;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56858.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 990S-0121825.  
PR 05-MAR-1999; 990S-0123180.  
PR 09-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130449.  
PR 23-APR-1999; 990S-0130510.  
PR 23-APR-1999; 990S-0130891.  
PR 28-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 30-APR-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 07-MAY-1999; 990S-0132487.  
PR 11-MAY-1999; 990S-0132863.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.

PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 17-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142820.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144684.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146389.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147021.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147792.

PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158923.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

48.4%; Score 46.5; DB 21; Length 252;

Best Local Similarity 43.5%; Pred. No. 18;  
Matches 10; Conservative 1; Mismatches 3; Indels 9; Gaps 1;  
QY 2 RLGFSRLP-----HTGCG 15  
I: | | | | | | | | | |  
Db 61 RVSFRRPGSVSLQRHFLGCG 83  
RESULT 15  
AAC24721  
ID AAC24721 standard; Protein; 256 AA.  
XX  
AC AAC24721;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 28504.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 200EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134421.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.





Tue May 20 18:04:53 2003

us-09-869-155-14.rag

Page 15

Search completed: May 20, 2003, 16:14:38  
Job time : 19.4945 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:11:22 ; Search time 6.91209 Seconds

(without alignments)  
236.439 Million cell updates/sec

Title: US-09-869-155-14

Perfect score: 96  
Sequence: 1 KRLGFSRLPHFTGCGGL 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46.5	48.4	253	2 D71446	hypothetical prote
2	45	46.9	290	2 C71701	branched-chain ami
3	43.5	45.3	438	2 T52082	alpha galactosyltr
4	43	44.8	206	2 AG0692	conserved hypothet
5	42	43.8	255	2 F83058	probable ATP-bind
6	42	43.8	268	2 T46902	hypothetical prote
7	42	43.8	325	2 T02117	hypothetical prote
8	42	43.8	472	2 T48074	hypothetical prote
9	42	43.8	1512	2 G85090	hypothetical prote
10	42	43.8	1534	2 S59604	hypothetical prote
11	42	43.8	3097	2 T28635	DNA (cytosine-5-)-
12	41.5	43.2	376	2 T49735	glutamate synthase
13	41.5	43.2	581	2 T05232	hepatocyte nuclear
14	41	42.7	96	2 T17865	hypothetical prote
15	41	42.7	178	2 T38624	hypothetical prote
16	41	42.7	211	2 A82915	hypothetical prote
17	41	42.7	333	2 S77222	ribosomal protein
18	41	42.7	431	2 AH0860	hypothetical prote
19	41	42.7	433	2 E65060	probable RNA methyl
20	41	42.7	433	2 F85929	ycga protein - Esc
21	41	42.7	457	2 T33244	probable enzyme y9
22	41	42.7	591	2 AC3528	extracellular prote
23	41	42.7	729	1 FOLJSA	gag polyprotein -
24	41	42.7	758	1 FOLJSA	probable sugar tra
25	41	42.7	855	2 JC7731	membrane-bound arg
26	41	42.7	855	2 JC7775	membrane type-seri
27	41	42.7	958	2 A70634	probable mmp11 pro
28	41	42.7	1528	2 D85912	hypothetical prote
29	41	42.7	1569	2 A65044	hypothetical prote

30	41	42.7	1571	2 C91068	hypothetical prote
31	40.5	42.2	212	2 S73991	ribosomal protein
32	40	41.7	82	2 JS0668	myomy protein - ra
33	40	41.7	88	2 A10041	hypothetical prote
34	40	41.7	133	2 G85643	hypothetical prote
35	40	41.7	133	2 E90783	hypothetical prote
36	40	41.7	172	2 E82464	probable acetyltra
37	40	41.7	324	2 JC0163	glucuronyltransfer
38	40	41.7	342	2 B70100	beta-N-acetylhexos
39	40	41.7	382	2 G90751	probable transport
40	40	41.7	382	2 E85615	probable transport
41	40	41.7	382	2 A64829	membrane protein y
42	40	41.7	473	2 S04113	nitrogenase (EC 1.
43	40	41.7	474	2 C35405	nitrogenase (EC 1.
44	40	41.7	528	2 B88989	protein F02C9.2 [1
45	40	41.7	564	2 F97668	probable electron

## ALIGNMENTS

## RESULT 1

D71446  
hypothetical protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

A:Variety: columbia

C>Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998

C:Accession: D71446

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D

P.; Wedler, H.; Wedler, E.; Wambutt, R.; Wellenreger, T.; Pohl, T.M.; Terry, N.; G

avenagh, T.; Hempel, S.; Kotter, P.; Eutlian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B

Nature 391, 485-488, 1998

A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen

erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;

C.; Chalwatzis, N.

A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t

A:Reference number: A71400; MUID:98121113; PMID:9461215

A:Accession: D71446

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-253 <BEV>

A:Cross-references: GB:T97343; NID:92245073; PID:e327075; PID:92245123

C:Genetics:

A:Map position: 4COP9-4G3845

Query Match 48.4%; Score 46.5; DB 2; Length 253;

Best Local Similarity 43.5%; Pred. No. 3.8;

Matches 10; Conservative 1; Mismatches 3; Indels 9; Gaps 1;

QY 2 RLGFSRLP-----HFTGCG 15

DB 127 RVSFRLPGSVSLQRRHFLGCG 149

## RESULT 2

C71701  
branched-chain amino acid aminotransferase (liver) RP428 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii

C>Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000

C:Accession: C71701

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark

Nature 396, 133-140, 1998

A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A:Reference number: A71630; MUID:99039499; PMID:9823893

A:Accession: C71701

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-290 <AND>

A:Cross-references: GB:AJ35271; GB:AJ235269; NID:g3868717; PIDN:CAA14885.1; PID:g386

A:Experimental source: strain Madrid E

C:Genetics:

A:Gene: lve; RP428

C:Superfamily: branched-chain-amino-acid aminotransferase

Query Match 46.9%; Score 45; DB 2; Length 290;  
 Best Local Similarity 50.0%; Pred. No. 7.7;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 KRUGSRPHFTGC 14  
 :|||:::|||||  
 Db 234 ERLKLAQIEHFTGC 247

## RESULT 3

T52082  
 alpha galactosyltransferase (EC 2.4.1.-) [imported] - Trigonella foenum-graecum (fragmen  
 C:Species: Trigonella foenum-graecum  
 C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
 C:Accession: T52082  
 R:Edwards, M.E.; Dickson, C.A.; Chengappa, S.; Sidebottom, C.M.; Gidley, M.J.; Reid, G.  
 Plant J. 19, 691-697, 1999  
 A:Title: Molecular characterisation of a membrane-bound galactosyltransferase of plant  
 A:Reference number: Z25943  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Accession: T52082  
 A:Molecule type: mRNA  
 A:Residues: 1-438 <EDW>  
 A:Cross-references: EMBL:AJ245478; PIDN:CAB52246.1  
 C:Genetics:  
 A:Gene: galtran  
 C:Function:  
 A:Pathway: galactomannan biosynthesis  
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 45.3%; Score 43.5; DB 2; Length 438;  
 Best Local Similarity 58.8%; Pred. No. 21;  
 Matches 10; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 1 KRUGSRP---HFTGC 14  
 :|||:::|||||  
 Db 363 KNLGDMKRPRTHTFTGC 379

## RESULT 4

AG0692  
 conserved hypothetical protein STY1667 [imported] - Salmonella enterica subsp. enterica  
 C:Species: Salmonella enterica subsp. enterica serovar Typh  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AG0692  
 R:Parhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 Th. T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AG0692  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-206 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD01912.1; PID:q16502754; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY1667  
 C:Superfamily: hypothetical protein H11687

Query Match 44.8%; Score 43; DB 2; Length 206;  
 Best Local Similarity 42.9%; Pred. No. 12;  
 Matches 9; Conservative 2; Mismatches 4; Indels 6; Gaps 1;

QY 1 KRUGF-----SRUPHFTGC 15  
 :|||:::|||||  
 Db 185 KRAGLYAESLPADLPHTTACG 205

## RESULT 5

F83058  
 Probable ATP-binding component of ABC transporter PA4706 [imported] - Pseudomonas aer  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: F83058  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
 Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; L  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
 A:Reference number: AB2950; PMID:20437337; PMID:10984043  
 A:Accession: F83058  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-255 <STO>  
 A:Cross-references: GB:AE004884; GB:AE004091; NID:g9950954; PIDN:AAG08092.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA4706

Query Match 43.8%; Score 42; DB 2; Length 255;  
 Best Local Similarity 63.6%; Pred. No. 22;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 LGFSRLPHFTG 13  
 :|||:::|||||  
 Db 96 VGFGRPLPHSSG 106

## RESULT 6

T46902  
 hypothetical protein DKFP761E10121.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 17-Mar-2000  
 C:Accession: T46902  
 R:Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Well, B.; Wilmann, S.  
 Submitted to the Protein Sequence Database, February 2000  
 A:Reference number: Z24135  
 A:Accession: T46902  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-268 <AA>  
 A:Cross-references: EMBL:AL157417  
 A:Experimental source: adult amygdala; clone DKFP761E10121  
 C:Genetics:  
 A:Note: DKFP761E10121.1

Query Match 43.8%; Score 42; DB 2; Length 268;  
 Best Local Similarity 50.0%; Pred. No. 23;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 RLGSRLPHFTGC 15  
 :|||:::|||||  
 Db 216 KVGFRLPHGWTGC 229

## RESULT 7

T02117  
 hypothetical protein At2g41020 [imported] - Arabidopsis thaliana  
 N:Alternate names: hypothetical protein T3K9.21  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 16-Feb-2001  
 C:Accession: T02117; G84836  
 R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke  
 submitted to the EMBL Data Library, February 1999  
 A:Description: Arabidopsis thaliana chromosome II BAC T3K9 genomic sequence.  
 A:Reference number: Z14570  
 A:Accession: T02117  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-325 <ROU>  
 A:Cross-references: EMBL:AC004261; NID:g3402695; PID:g3402715  
 A:Experimental source: cultivar Columbia

R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Weiss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A64420; MUID:20083487; PMID:10617197

A:Accession: G84836

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-325 <STO>

A:Cross-references: GB:AE002093; NID:93402715; PIDN:AA012009.1; GSPDB:GN00139

C:Genetics:

A:Gene: T3K9.21; At2g41020

A:Map position: 2

A:Introns: 44/3; 56/2; 83/3; 111/1; 173/3; 198/3; 252/1

C:Superfamily: Arabidopsis thaliana hypothetical protein At2g41020; WW repeat homology F:192-220/Domain: WW repeat homology <MW1>

F:237-275/Domain: WW repeat homology <MW2>

Query Match 43.8%; Score 42; DB 2; Length 325;  
Best Local Similarity 63.6%; Pred. No. 28;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 6 SRLPHFTGCGG 16  
|:|:| |111|  
Db 298 SOLPRCSGCGG 308

## RESULT 8

T48074  
Hypothetical protein F26K9.280 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T48074  
R:Blecker, H.; Menez, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: 22465  
A:Accession: T48074  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-472 <BLO>  
A:Cross-references: EMBL:AL162651  
A:Experimental source: cultivar Columbia; BAC clone F26K9  
C:Genetics:  
A:Map position: 3  
A:Introns: 85/1; 208/3; 309/1  
A:Note: F26K9.280

Query Match 43.8%; Score 42; DB 2; Length 472;  
Best Local Similarity 53.3%; Pred. No. 40;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 KRLGFSRLPHFTGCGG 15  
|:|:|:| |111|  
Db 269 KRLGSGPVSYGCGG 283

## RESULT 9

G85090  
Hypothetical protein AT4g08990 (imported) - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: G85090  
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A65001; MUID:20083488; PMID:10617198  
A:Accession: G85090  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1512 <STO>  
A:Cross-references: GB:NC\_001268; NID:97267541; PIDN:CAE78023.1; GSPDB:GN00140  
C:Genetics:

A:Gene: AT4g08990  
A:Map position: 4

Query Match 43.8%; Score 42; DB 2; Length 1512;  
Best Local Similarity 47.1%; Pred. No. 1,2e+02;  
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 KRLGFSRLPHFTGCGG 17  
|:|:|:| |1111|  
Db 1067 KEMRLATLDFAGCGG 1083

## RESULT 10

S59604  
DNA (cytosine-5-)-methyltransferase (EC 2.1.1.37) - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 30-Nov-1995 #sequence\_revision 19-Jan-1996 #text\_change 21-Jul-2000  
C:Accession: S59604  
R:Finnegan, E.J.; Dennis, E.S.  
Nucleic Acids Res. 21, 2383-2388, 1993  
A:Title: Isolation and identification by sequence homology of a putative cytosine met  
A:Reference number: S35641; MUID:93281384; PMID:8385441  
A:Accession: S59604  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1534 <FIN>  
A:Cross-references: EMBL:L10692; NID:9304106; PIDN:AAA32829.1; PID:9304107  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993  
C:Keywords: methyltransferase; S-adenosylmethionine

Query Match 43.8%; Score 42; DB 2; Length 1534;  
Best Local Similarity 47.1%; Pred. No. 1.3e+02;  
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 KRLGFSRLPHFTGCGG 17  
|:|:|:| |1111|  
Db 1089 KEIRLATLDFAGCGG 1105

## RESULT 11

T28635  
glutamate synthase (NADH2) (EC 1.4.1.14) - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-Jun-2002  
C:Accession: T28635  
R:Cowan, G.M.  
submitted to the EMBL Data Library, April 1998  
A:Reference number: 220490  
A:Accession: T28635  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3097 <COW>  
A:Cross-references: EMBL:Y17045; NID:e1286063; PID:e1286064; PIDN:CAA76602.1  
C:Genetics: glts  
A:Gene: glts  
C:Keywords: oxidoreductase

Query Match 43.8%; Score 42; DB 2; Length 3097;  
Best Local Similarity 58.3%; Pred. No. 2.5e+02;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 4 GFSRLPHFTGCGG 15  
|:|:|:| |111|  
Db 992 GISTLPSTYKCGG 1003

## RESULT 12

I49735  
hepatocyte nuclear factor 3 forkhead homolog 8 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 24-Sep-1999  
C:Accession: I49735



GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:05:42 ; Search time 4.57692 Seconds

(without alignments)  
154.055 Million cell updates/sec

Title: US-09-869-155-14

Perfect score: 96  
Sequence: 1 KRLGSRPLPHTGCGGL 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.5	49.5	873	1 RX_DROME	Q9W2Q1 drosophila
2	45	46.9	290	1 ILVE_RICPR	O05970 rickettsia
3	44	45.8	933	1 PLD2_MOUSE	P97813 mus musculus
4	44	45.8	933	1 PLD2_MOUSE	P70498 rattus norv
5	43	44.8	206	1 RNF3_RAT	O82697 salmone
6	42	43.8	1534	1 DNMI_ARATH	P34861 arabidopsis
7	41.5	43.2	353	1 FXR1_MOUSE	O61080 mus musculu
8	41	42.7	338	1 HIS8_PPRFU	O81225 pyrococcus
9	41	42.7	430	1 RUMA_SALTI	O82426 salmone
10	41	42.7	432	1 RUMA_SALTY	O82426 salmone
11	41	42.7	432	1 RUMA_SALTY	O82426 salmone
12	41	42.7	432	1 RUMA_SALTY	O82426 salmone
13	41	42.7	432	1 RUMA_SALTY	O82426 salmone
14	41	42.7	432	1 RUMA_SALTY	O82426 salmone
15	41	42.7	432	1 RUMA_SALTY	O82426 salmone
16	40.5	42.2	212	1 RLA_MYCPN	P75579 mycoplasma
17	40	41.7	265	1 VNEP_AZOPA	O68953 azotobacter
18	40	41.7	279	1 VNEP_AZOSA	O68953 azotobacter
19	40	41.7	324	1 B3G2_RAT	O92437 rattus norv
20	40	41.7	382	1 YCAD_ECOLI	P15303 azotobacter
21	40	41.7	472	1 VNEP_AZOC	P15303 azotobacter
22	40	41.7	473	1 VNEP_AZOC	P15303 azotobacter
23	40	41.7	515	1 Y141_HUMAN	O14154 homo sapien
24	40	41.7	628	1 GIDA_NEIMA	O93441 neisseria m
25	40	41.7	628	1 GIDA_NEIMA	O93441 neisseria m
26	39	40.6	121	1 FLIO_ECOLI	P22886 escherichia
27	39	40.6	156	1 VEG6_HPV41	P27355 human papil
28	39	40.6	178	1 TRSF_DROER	O23335 drosophila
29	39	40.6	215	1 RLA_MYCBO	O06114 mycobacteri
30	39	40.6	223	1 RLA_MYCBO	O06045 mycobacteri
31	39	40.6	223	1 RLA_MYCBO	P95050 mycobacteri
32	39	40.6	231	1 RNC_SYNY3	P74368 synchocyst
33	39	40.6	302	1 TFPI_RAT	O02445 rattus norv

34	39	40.6	442	1 VAN1_CANAL	Q00314 candida alb
35	39	40.6	889	1 C122_HUMAN	O92966 homo sapien
36	38.5	40.1	488	1 ARI2_CAEEL	O22431 caenorhabdi
37	38	39.6	135	1 RT16_MOUSE	O92967 mus musculu
38	38	39.6	137	1 RT16_MOUSE	O92967 mus musculu
39	38	39.6	161	1 DUN_RHILLO	O92967 mus musculu
40	38	39.6	295	1 MTL5_MOUSE	O92967 mus musculu
41	38	39.6	303	1 YE15_YEAST	P40100 saccharomyc
42	38	39.6	304	1 MTH5_HAEIN	P45000 haemophilus
43	38	39.6	330	1 MTP2_NEIGO	P08455 neisseria g
44	38	39.6	374	1 MTF7_NEIGO	O92966 neisseria g
45	38	39.6	412	1 PEPT_PASWU	O92966 pasteurrella

## ALIGNMENTS

RESULT 1	ID	RX_DROME	STANDARD	PRT	873 AA.
AC	Q9W2Q1	O46035			
DT	15-OCT-2001	(Rel. 40, Created)			
DT	15-JUN-2002	(Rel. 41, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Retinal homeobox protein Rx (DRX1) (DRX).				
GN	RX OR CG10052.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;				
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;				
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Cancon-S; TISSUE=Embryo;				
RX	MEDLINE=98151514; PubMed=9482887;				
RA	Eggert T., Hauck B., Hildebrandt N., Gehring W.J., Walldorf U.;				
RT	Isolation of a Drosophila homolog of the vertebrate homeobox gene Rx				
RT	and its possible role in brain and eye development."				
RT	Proc. Natl. Acad. Sci. U.S.A. 95:2343-2348(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Berkley;				
RX	MEDLINE=20196006; PubMed=10731132;				
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	Georgie R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,				
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
RA	Abriil J.F., Aghayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,				
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,				
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Durbin K.J., Evangelista C.C., Ferraz C., Fertler S., Fleischmann W.,				
RA	Foster G., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
RA	Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,				
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,				
RA	Jalali B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,				
RA	Merklou G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,				
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,				
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,				
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,				
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,				
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,				
RA	Spieler E., Spradling A.C., Stapleton M., Strong R., Sun E.,				

RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster.",  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP REVISIONS.  
 RC STRAIN-Berkeley:  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.E., Bergman C.M., Berman B.P., Carlson J.W., Celisner S.E.,  
 RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,  
 RA Harris N.L., Krommiller B., Marshall B., Millburn G.H., Richter J.,  
 RA Russo S., Searle S.M.J., Smith E., Shu S., Smutnak F.,  
 RA Whitfield E.J., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Kungall C.J., Lewis S.E.,  
 RT "Annotation of Drosophila melanogaster genome.",  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RX MEDLINE-97320497; PubMed-9177348;  
 RA Natchers P.H., Grinberg A., Mahon K.A., Jamrich M.;  
 RT "The Rx homeobox gene is essential for vertebrate eye development.";  
 CC Nature 387:603-607(1997).  
 CC -1- FUNCTION: Appears to function in brain development.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- DEVELOPMENTAL STAGE: Expressed in the precephalic region and in  
 CC the clypeolabrum from stage 8 on and later in the brain and the  
 CC central nervous system.  
 CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
 CC "BICOID" SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ223300; CA11241.1; ALT\_INT.  
 DR EMBL: AE003452; AAF46639.2; ALT\_SEQ.  
 DR TRANSFAC: T03511; -  
 DR FlyBase: FBgn0020617; Rx.  
 DR InterPro: IPR003654; Homeo\_OAR.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox; 1.  
 DR ProDom: PD000010; Homeobox; 1.  
 DR SMART: SM00389; HOX; 1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 DR PROSITE: PS50803; OAR; 1.  
 DR Homeobox: DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 KM  
 FT DOMAIN 116 123 OCTAPEPTIDE MOTIF.  
 FT DNA\_BIND 526 586 HOMEBOX.  
 FT DOMAIN 849 862 OAR.  
 FT DOMAIN 855 859 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 329 333 POLY-PRO.  
 FT DOMAIN 366 369 POLY-ASP.  
 FT DOMAIN 437 453 POLY-GLN.  
 FT DOMAIN 670 673 POLY-PRO.  
 FT DOMAIN 683 691 POLY-PRO.  
 FT DOMAIN 753 763 POLY-SER.  
 FT DOMAIN 828 834 POLY-PRO.  
 FT CONFLICT 4 4 S -> P (IN REF. 1).

FT CONFLICT 136 136 R -> W (IN REF. 1).  
 FT CONFLICT 384 384 S -> T (IN REF. 1).  
 FT CONFLICT 645 672 PLSLAPGNLTMSLSLAAHGHHANGPPP -> QGARKSDH  
 FT CONFLICT 768 768 EOSGGGPPPCQWMAA (IN REF. 1).  
 FT CONFLICT 768 768 G -> L (IN REF. 1).  
 SO SEQUENCE 873 AA: 92897 MW: 87856AD693F6710 CRC64;  
 Query Match 49.5%; Score 47.5; DB 1; Length 873;  
 Best Local Similarity 58.8%; Pred. No. 3.4;  
 Matches 10; Conservative 2; Mismatches 2; Indels 3; Gaps 1;  
 QY 2 RLG---FSRLPHFTGCG 15  
 DB 593 RLGRLHTQPLPHRLGCG 609  
 RESULT 2  
 ILVE\_RICPR  
 ID ILVE\_RICPR STANDARD; PRT; 290 AA.  
 AC 005970;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable branched-chain amino acid aminotransferase (EC 2.6.1.42)  
 DE (BCAT).  
 GN ILVE OR RP428.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiense; Rickettsia.  
 OX NCBI\_TaxId=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Madrid E;  
 RX MEDLINE-97419517; PubMed-9274032;  
 RA Andersson J.O., Andersson S.G.E.;  
 RT Intracellular parasite Rickettsia prowazekii as inferred from an  
 RL analysis of 52015 bp nucleotide sequence.";  
 RN Microbiology 143:2783-2795(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Madrid E;  
 RX MEDLINE-9823893; PubMed-9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Sichteritz-Ponten T., Almark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria".  
 RL Nature 396:133-140(1998).  
 CC -1- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).  
 CC -1- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-  
 CC oxopentanoate + L-glutamate.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- PATHWAY: Valine and isoleucine biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y11777; CA172450.1; -  
 DR EMBL: AJ235271; CA14885.1; -  
 DR HSSP: P00510; IA3G.  
 DR InterPro: IPR001544; Aminotran\_4.  
 DR Pfam: PF01063; aminotran\_4; 1.  
 DR ProDom: PD001961; Aminotran\_4; 1.  
 DR PROSITE: PS00770; AA-TRANSFER\_CLASS\_4; 1.  
 DR Transferase: Aminotransferase; Branched-chain amino acid biosynthesis;  
 KW



KW Pyridoxal phosphate; Complete proteome.  
 FT BINDING 155 155 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SO SEQUENCE 290 AA; 33068 MW; 98374E325350763D CRC64;  
 Query Match 46.9%; Score 45; DB 1; Length 290;  
 Best Local Similarity 50.0%; Pred. No. 3;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 KRIFSRLPHFTGCG 14  
 DB 234 ERLKAOIEHFTGC 247

RESULT 3  
 PLD2\_MOUSE STANDARD; PRT; 933 AA.  
 AC P97813;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phospholipase D2 (EC 3.1.4.4) (PLD 2) (Choline phosphatase 2)  
 GN Phosphatidylcholine-hydrolyzing phospholipase D2 (PLD1C) (mPLD2).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Embryo, and Neonatal brain;  
 RA MEDLINE=97199446; PubMed=9395408;  
 RA Colley W.C., Sung T.-C., Roll R., Jenco J.M., Hammond S.M.,  
 RA Altschuler Y.M., Bar-Sagi D., Morris A.J., Frohman M.A.;  
 RT "Phospholipase D2, a distinct phospholipase D isoform with novel  
 RT regulatory properties that provokes cytoskeletal reorganization.";  
 RT Curr. Biol. 7:191-201(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RA MEDLINE=98228223; PubMed=9560313;  
 RA Redina O.E., Frohman M.A.;  
 RT "Organization and alternative splicing of the murine phospholipase D2  
 RT gene.";  
 RT Biochem. J. 331:845-851(1998).  
 RN [3]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=97439716; PubMed=9307024;  
 RA Colley W.C., Altschuler Y.M., Sue-Ling C.K., Copeland N.G.,  
 RA Gilbert D.J., Jenkins N.A., Branch K.D., Tsifka S.E., Bollag R.J.,  
 RA Bollag W.B., Frohman M.A.;  
 RT "Cloning and expression analysis of murine phospholipase D1.";  
 RT Biochem. J. 326:745-753(1997).  
 CC -1- FUNCTION: MAY HAVE A ROLE IN SIGNAL-INDUCED CYTOSKELETAL  
 CC REGULATION AND/OR ENDOCYTOSIS.  
 CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a  
 CC phosphatidate.  
 CC -1- ENZYME REGULATION: STIMULATED BY PHOSPHATIDYLINOSITOL 4,5-  
 CC BISPHOSPHATE. IS NOT RESPONSIVE TO ADP-RIBOSYLATION FACTOR-1 (ARF-  
 CC 1), NOR TO GTP-BINDING PROTEINS. RHO A.  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS. ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: UBQUITOUS. HIGHEST LEVELS IN BRAIN AND LUNG.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT HIGH LEVELS IN THE HIPPOCAMPUS  
 CC ALSO IN VENTRICULAR NEURAL CELLS AS WELL AS DIFFERENTIATING  
 CC NEURONS OUTSIDE OF THE VENTRICULAR REGION. EXPRESSED DURING  
 CC DEVELOPMENT IN LOWER LEVELS IN MESENCHYMAL CELLS DERIVED FROM THE  
 CC NEURAL CREST THAT ARE DESTINED TO FORM BONES OF THE SKULL.  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U87557; AAC53173.1; -  
 CC DR EMBL: AF052294; AAC24519.1; -  
 CC DR EMBL: AF052291; AAC24519.1; JOINED.  
 CC DR EMBL: AF052293; AAC24519.1; JOINED.  
 CC DR EMBL: AF052292; AAC24519.1; JOINED.  
 CC MGD: MGI:892877; Pld2.  
 CC DR InterPro: IPR001849; PH.  
 CC DR InterPro: IPR001736; PLD.  
 CC DR InterPro: IPR001683; PX.  
 CC DR Pfam: PF00169; PH; 1.  
 CC DR Pfam: PF00614; PLDC; 2.  
 CC DR Pfam: PF00787; PX; 1.  
 CC DR SMART: SM00233; PH; 1.  
 CC DR SMART: SM00155; PLDC; 2.  
 CC DR SMART: SM00312; PX; 1.  
 CC DR PROSITE: PS50003; PH\_DOMAIN; FALSE\_NEG.  
 CC DR PROSITE: PS50035; PLD; 2.  
 CC DR PROSITE: PS50195; PX; 1.  
 CC DR Hydrolase; Lipid degradation; Membrane; Alternative splicing; Repeat.  
 CC FT DOMAIN 65 195 PX.  
 CC FT DOMAIN 203 311 PH.  
 CC FT DOMAIN 437 464 PLD PHOSPHODIESTERASE 1.  
 CC FT DOMAIN 751 778 PLD PHOSPHODIESTERASE 2.  
 CC FT DOMAIN 441 788 CATALYTIC.  
 CC SO SEQUENCE 933 AA; 106167 MW; BADE1EDFEAC9ED CRC64;

Query Match 45.8%; Score 44; DB 1; Length 933;  
 Best Local Similarity 60.0%; Pred. No. 14;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 KRIFSRLPHFTGCG 15  
 DB 211 KRSGHVRVPGFTFCG 225

RESULT 4  
 PLD2\_RAT STANDARD; PRT; 933 AA.  
 ID PLD2\_RAT  
 AC P70498; 008768;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phospholipase D2 (EC 3.1.4.4) (PLD 2) (Choline phosphatase 2)  
 GN Phosphatidylcholine-hydrolyzing phospholipase D2 (PLD1C) (rPLD2).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98194218; PubMed=9533024;  
 RA Nakashima S., Matsuda Y., Akao Y., Yoshimura S.-I., Sakai H.,  
 RA Hayakawa K., Andoh M., Nozawa Y.;  
 RT "Molecular cloning and chromosome mapping of rat phospholipase D  
 RT genes, Pld1a, Pld1b and Pld2.";  
 RT Cytogenet. Cell Genet. 79:109-113(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
 RC TISSUE-Brain;  
 RX MEDLINE=97269050; PubMed=9111050;  
 RA Kodaki T., Yamashita S.;  
 RT "Cloning, expression, and characterization of a novel phospholipase D  
 RT complementary DNA from rat brain.";

```

RL J. Biol. Chem. 272:11408-11413(1997).
RN [3]
RP SEQUENCE OF 445-535 FROM N.A.
RX TISSUE-Glial cell;
RX MEDLINE=96354814; PubMed=8753790;
RA Yoshimura S.-I., Nakashima S., Ohguchi K., Sakai H., Shinoda J.,
RA Sakai N., Nozawa Y.;
RT "Differential mRNA expression of phospholipase D (PLD) isozymes
RT during CAMP-induced differentiation in C6 glioma cells.";
RL Biochem. Biophys. Res. Commun. 225:494-499(1996).
CC -1- FUNCTION: MAY HAVE A ROLE IN SIGNAL-INDUCED CYTOSKELETAL
CC REGULATION AND/OR ENDOCYTOSIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a
CC phosphatidate.
CC -1- ENZYME REGULATION: STIMULATED BY PHOSPHATIDYLINOSITOL 4,5-
CC BISPHOSPHATE, AND PHOSPHATIDYLETHANOLAMINE. INHIBITED BY
CC PHOSPHATIDYLSELINE AND BY OLEATE. IS NOT RESPONSIVE TO ADP-
CC RIBOSYLATION FACTOR-1 (ARF-1), NOR TO GTP-BINDING PROTEINS: RHO A.
CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG, HEART, KIDNEY,
CC STOMACH, SMALL INTESTINE, COLON, AND TESTIS, AND AT A MUCH LOWER
CC LEVELS IN THYMUS, LIVER AND MUSCLE.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB003172; BAA24078.1; -
DR EMBL: D88672; BAA19882.1; -
DR InterPro: IPR001849; PH.
DR InterPro: IPR001736; PLD.
DR InterPro: IPR001683; PX.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00614; PLDc; 2.
DR Pfam: PF00787; PX; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00153; PLDc; 2.
DR SMART: SM00312; PX; 1.
DR PROSITE: PS50003; PH_DOMAIN; FALSE_NEG.
DR PROSITE: PS50035; PLD; 2.
DR PROSITE: PS50195; PX; 1.
DR Hydrolyase; Lipid degradation; Membrane; Repeat.
KW DOMAIN
FT 65 195
FT DOMAIN
FT 203 311
FT DOMAIN
FT 437 464
FT DOMAIN
FT 751 778
FT DOMAIN
FT 441 768
FT CONFLICT
FT 26 26
FT CONFLICT
FT 125 125
FT CONFLICT
FT 599 599
FT CONFLICT
FT 792 792
FT CONFLICT
FT 817 818
FT CONFLICT
FT 919 924
SQ SEQUENCE 933 AA; 106036 MW; D430843B4D541EEA CRC64;

```

Query Match 45.8%; Score 44; DB 1; Length 933;  
 Best Local Similarity 60.0%; Pred. No. 14;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 KRIGFSLPHFTGCG 15  
 11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
 DB 211 KRSGGHRVPGFTCGC 225

RESULT 5

```

RNFG_SALTI
ID RNFG_SALTI STANDARD; PRT; 206 AA.
AC 082607;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Electron transport complex protein rnfG.
DE RNFG OR STY1667.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21534947; PubMed=11677608;
RX STRAIN=C18;
RA Barthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Kung'u K.L., Bentley S.D., Holden M.T.G., Sebahina M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi C18."
RL Nature 413:848-852(2001).
CC -1- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -1- SUBCELLULAR LOCATION: Inner membrane (Potential).
CC -1- SIMILARITY: BELONGS TO THE RNFG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL627271; CAD01912.1; -
DR Electron transport; Transmembrane; Inner membrane; Complete proteome.
KW TRANSMEM 10 26
KW POTENTIAL.
SQ SEQUENCE 206 AA; 22252 MW; D5FBA792801B34A CRC64;

```

Query Match 44.8%; Score 43; DB 1; Length 206;  
 Best Local Similarity 42.9%; Pred. No. 4.6;  
 Matches 9; Conservative 2; Mismatches 4; Indels 6; Gaps 1;

OY 1 KRIGFSLPHFTGCG 15  
 11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
 DB 185 KRAGIYAESLPADLPHTAGC 205

RESULT 6  
 DNML\_ARATH STANDARD; PRT; 1534 AA.  
 ID DNML\_ARATH  
 AC P34881;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA (cytosine-5)-methyltransferase Ath1 (EC 2.1.1.37) (DNA  
 DE methyltransferase Ath1) (DNA Metase Ath1) (M.Ath1).  
 GN ATH1 OR AT5G49160 OR K2IP3.3.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;  
 RX MEDLINE=93281384; PubMed=8389441;  
 RA Finnegan E.J., Dennis E.S.;  
 RT "Isolation and identification by sequence homology of a putative  
 RL cytosine methyltransferase from Arabidopsis thaliana.";  
 RL Nucleic Acids Res. 21:2383-2388(1993).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=99156233; PubMed=10048488;  
 RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.  
 RT Sequence features of the regions of 1,081,958 bp covered by seventeen  
 RT physically assigned P1 and TAC clones.";  
 RL DNA Res. 5:379-391(1998).  
 CC -1- FUNCTION: Methylates CPG residues.  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA = S-adenosyl-L-  
 CC homocysteine + DNA containing 5-methylcytosine.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 BAH DOMAINS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 DR EMBL: L10692; AAA32829.1; -  
 DR HSSP: O14717; 1G55  
 DR REBASE: 2839; M.Ath1.  
 DR InterPro: IPR001025; BAH.  
 DR InterPro: IPR001525; C5\_DNA\_meth.  
 DR Pfam: PF00145; DNA\_methylase; 1.  
 DR Pfam: PF01426; BAH; 2.  
 DR PRINTS: PR00105; C5METTRFRASE.  
 DR SMART: SM00439; BAH; 2.  
 DR TIGRfam: TIGR00675; dcm; 1.  
 DR PROSITE: PS00094; C5\_MTASE\_1; 1.  
 DR PROSITE: PS00095; C5\_MTASE\_2; 1.  
 KW Transferrase; Methyltransferase; DNA-binding; Nuclear protein; Repeat.  
 FT DOMAIN 735 869 BAH 1.  
 FT DOMAIN 909 1049 BAH 2.  
 FT ACT\_SITE 1198 1198 BY SIMILARITY.  
 FT SEQUENCE 1534 AA; 172430 MW; 23FC944AA0704C5A CRC64;  
 SQ  
 Query Match 43.8%; Score 42; DB 1; Length 1534;  
 Best Local Similarity 47.1%; Pred. No. 50;  
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 KRLGFSRLPHFTGCGGL 17  
 DB 1089 KEIRLATLDFRAGCGGL 1105

RESULT 7  
 FXFL\_MOUSE STANDARD: PRT; 353 AA.  
 AC 061080; 061661;  
 DT 01-NOV-1997 (Rel. 35; Last sequence update)  
 DT 16-OCT-2001 (Rel. 40; Last annotation update)  
 DE Forkhead box protein F1 (Forkhead-related protein FKHL5) (Forkhead-  
 DE related transcription factor 1) (FREC-1) (hepatocyte nuclear factor 3  
 DE forkhead homolog 8) (HFH-8).  
 GN FOXF1 OR FOXF1A OR FKHL5 OR FREAC1 OR HFH8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=96224034; PubMed=8626802;  
 RA Helliqvist M., Mahlapuu M., Samuelsson L., Enerbaeck S., Carlsson P.;  
 RT "Differential activation of lung-specific genes by two forkhead  
 RT proteins, FREAC-1 and FREAC-2.";  
 RL J. Biol. Chem. 271:4482-4490(1996).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/c;  
 RX MEDLINE=95046902; PubMed=7958446;  
 RA Clevidence D.E., Overdier D.G., Peterson R.S., Porcella A., Ye H.,  
 RA Paulson K.E., Costa R.H.;  
 RT "Members of the HNF-3/forkhead family of transcription factors  
 RT exhibit distinct cellular expression patterns in lung and regulate  
 RT the surfactant protein B promoter.";  
 RL Dev. Biol. 166:195-209(1994).  
 CC -1- FUNCTION: PROBABLE TRANSCRIPTION ACTIVATOR FOR A NUMBER OF LUNG-  
 CC SPECIFIC GENES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: LUNG AND PLACENTA.  
 CC -1- DOMAIN: ACTIVATION DOMAINS C-TERMINAL OF (AND DISTINCT FROM) THE  
 CC FORKHEAD DOMAINS ARE NECESSARY FOR TRANSCRIPTIONAL ACTIVATION (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 DR EMBL: U42596; AAC52445.1; -  
 DR EMBL: L35949; AAA64885.1; ALT\_INIT.  
 DR HSSP: 063245; 2HFH.  
 DR TRANSFAC: T02461;  
 DR MGD: MGI:1347470; Foxf1a.  
 DR InterPro: IPR001766; TF\_Fork\_head.  
 DR Pfam: PF00250; Fork\_head; 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR ProDom: PD000425; TF\_Fork\_head; 1.  
 DR SMART: SM00339; FH; 1.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 KW DNA-binding; Nuclear protein; Transcription regulation; Activator.  
 FT DNA\_BIND 22 113 FORK-HEAD.  
 FT CONFLICT 7 19 GPTKAKTNAGVR -> PHQGEQRRRA (IN  
 FT REF. 2).  
 FT CONFLICT 207 247 AGEYPHDSVPASPLPAGAGVMEPAHYSSSAAMP  
 FT -> GRGVAPAPQLGAFRTAPAPAEWSRTPFPALQPG  
 FT RP (IN REF. 2).  
 FT SEQUENCE 353 AA; 37798 MW; 046AED08D1765A69 CRC64;  
 SQ  
 Query Match 43.2%; Score 41.5; DB 1; Length 353;  
 Best Local Similarity 55.6%; Pred. No. 14;  
 Matches 10; Conservative 1; Mismatches 4; Indels 3; Gaps 1;

OY 3 LGFSRLPH---FTGCGGL 17  
 DB 136 LGFNHLPTDYGFGSGGL 153

RESULT 8  
 HIS8\_PYRFU STANDARD: PRT; 338 AA.  
 AC 08TH25;  
 DT 15-JUN-2002 (Rel. 41; Last sequence update)

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Histidinol-phosphate aminotransferase (EC 2.6.1.9) (Imidazole acetol-
GN Hisc OR PFI665.
OS Pyrococcus furiosus.
OC Archaea: Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VCL / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-histidinol-phosphate + 2-oxoglutarate -> 3-
CC (Imidazol-4-yl)-2-oxopropyl phosphate + L-glutamate.
CC -1- COFACTOR: Pyridoxal phosphate (By similarity).
CC -1- PATHWAY: Histidine biosynthesis; seventh step.
CC -1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES. HISTIDINOL-PHOSPHATE AMINOTRANSFERASES
CC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: AE010265; AAL81789.1; -
DR PROSITE: PS00599; AA_TRANSF_2; 1.
KW Histidine biosynthesis; Transferase; Aminotransferase;
KW Pyridoxal phosphate; Complete proteome.
FT BINDING 204 204 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 338 AA; 39086 MW; A1781B9B375016C CRC64;

Query Match
Best Local Similarity 42.7%; Score 41; DB 1; Length 338;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KRUGSRLPHFT 12
   11: 1 1 1 1 1 1
Db 45 KRIEFNRPYHIT 56

RESULT 9
RUMA_SALT1
ID RUMA_SALT1 STANDARD: PRT; 430 AA.
AC Q82446;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 23S rRNA (Uracil-5-)-methyltransferase rumA (EC 2.1.1.-) (23S rRNA(M-
DE 5-01939)-methyltransferase).
GN RUMA OR STY3095.
OS Salmonella typhi.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21534947; PubMed=11677608;
RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
RA Crocin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella

```

```

RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
CC 1939 (M-5-01939) in 23S rRNA (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-
CC homocysteine + rRNA containing thymine.
CC -1- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. RUMA
CC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: AL62726; CAD06071.1; -
DR TIGRfams: TIGR00479; y9CA.1.
DR PROSITE: PS01230; TRMA_1; 1.
DR PROSITE: PS01231; TRMA_2; 1.
KW Transferase; Methyltransferase; rRNA processing; Complete proteome.
FT INIT_MER 0 0 BY SIMILARITY.
FT ACT_SITE 367 367
SQ SEQUENCE 430 AA; 47745 MW; 53566AE51BD6376 CRC64;

Query Match
Best Local Similarity 42.7%; Score 41; DB 1; Length 430;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 RLPHFTGCGG 16
   1 1 1 1 1 1 1 1
Db 79 RCPHFVCGG 88

RESULT 10
RUMA_SALT1
ID RUMA_SALT1 STANDARD: PRT; 430 AA.
AC Q82ME1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 23S rRNA (Uracil-5-)-methyltransferase rumA (EC 2.1.1.-) (23S rRNA(M-
DE 5-01939)-methyltransferase).
GN RUMA OR STM2957.
OS Salmonella typhimurium.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC412 / ATCC 700720;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvoney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
CC 1939 (M-5-01939) in 23S rRNA (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-
CC homocysteine + rRNA containing thymine.
CC -1- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. RUMA
CC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: AE008835; AAL21837.1; -  
DR TRIGRAMS: TRIGR00479; ygcA; 1.  
DR PROSITE: PS01230; TRNA\_1; 1.  
DR PROSITE: PS01231; TRNA\_2; 1.  
DR StyGene: SG77777; ruma.  
KW Transferase; Methyltransferase; RNA processing; Complete proteome.  
FT INIT MET 0 BY SIMILARITY.  
FT ACT\_SITE 387 387 BY SIMILARITY.  
SQ SEQUENCE 430 AA; 47636 MW; 4C415A92515D28DF CRC64;  
  
Query Match 42.7%; Score 41; DB 1; Length 430;  
Best Local Similarity 70.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 7 RLPHFTGCGG 16  
I I I I I  
Db 79 RCPHFGVCGG 88  
  
RESULT 11  
RUMA\_ECO57 STANDARD; PRT; 432 AA.  
ID RUMA\_ECO57  
AC Q8XED8;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 23S rRNA (Uracil-5'-methyltransferase ruma (EC 2.1.1.-) (23S rRNA(M-5-U1939)-methyltransferase).  
GN RUMA OR Z4100 OR ECS3645.  
OS Escherichia coli O157:H7.  
OC Escherichia: Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia  
OX NCBI\_TaxID=83334;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RL "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";  
RN Nature 409:529-533(2001).  
RT [2]  
RU SEQUENCE FROM N.A.  
RP STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Tida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position  
CC 1939 (M-5-U1939) in 23S rRNA (By similarity).  
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-  
CC homocysteine + rRNA containing thymine.  
CC -1- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. RUMA  
CC SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE005506; AAG57898.1; -  
DR EMBL: AP002562; BAB37068.1; -  
DR TRIGRAMS: TRIGR00479; ygcA; 1.  
DR PROSITE: PS01230; TRNA\_1; 1.  
DR PROSITE: PS01231; TRNA\_2; 1.  
KW Transferase; Methyltransferase; RNA processing; Complete proteome.  
FT INIT MET 0 BY SIMILARITY.  
FT ACT\_SITE 388 388 BY SIMILARITY.  
FT CONFLICT 88 88 G->S (IN REF. 2).  
SQ SEQUENCE 432 AA; 48024 MW; 9D7F8234E81F3ADE CRC64;  
  
Query Match 42.7%; Score 41; DB 1; Length 432;  
Best Local Similarity 70.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 7 RLPHFTGCGG 16  
I I I I I  
Db 79 RCPHFGVCGG 88  
  
RESULT 12  
RUMA\_ECOLI STANDARD; PRT; 432 AA.  
ID RUMA\_ECOLI  
AC P5135;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 23S rRNA (Uracil-5'-methyltransferase ruma (EC 2.1.1.-) (23S rRNA(M-5-U1939)-methyltransferase).  
GN RUMA OR B2785.  
OS Escherichia coli.  
OC Escherichia: Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia  
OX NCBI\_TaxID=562;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RL "The complete genome sequence of Escherichia coli K-12.";  
RN Science 277:1453-1474(1997).  
RT [2]  
RU SEQUENCE OF 1-10, AND CHARACTERIZATION.  
RP MEDLINE=21883956; PubMed=11779873;  
RX Agarwalla S., Kealey J.T., Santl D.V., Stroud R.M.;  
RT "Characterization of the 23 S ribosomal RNA m5U1939 methyltransferase  
from Escherichia coli.";  
RL J. Biol. Chem. 277:8835-8840(2002).  
CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position  
CC 1939 (M-5-U1939) in 23S rRNA.  
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-  
CC homocysteine + rRNA containing thymine.  
CC -1- MISCELLANEOUS: Two bound iron atoms are present in each monomer of  
CC ruma. However, the proposed mechanism does not involve metal ions  
CC in the reaction, so it seems unlikely that iron is required for  
CC the methyltransferase activity.  
CC -1- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. RUMA  
CC SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EcGene; EG11247; fuma.  
 DR InterPro: IPR002792; DUF90.  
 DR InterPro: IPR000051; SAM\_bind.  
 DR InterPro: IPR001566; TRMA\_1.  
 DR Pfam: PF01938; TRAM; 1.  
 DR TIGRFAMs: TIGR00479; y9ca.1.  
 DR PROSITE: PS01230; TRMA\_1; 1.  
 DR PROSITE: PS01231; TRMA\_2; 1.  
 KM Transferase; Methyltransferase; rRNA processing; Complete proteome.  
 FT INT\_MET 0  
 FT ACT\_SITE 388  
 SO SEQUENCE 432 AA; 47921 MW; DB8CB9F9E5FAB3 CRC64;  
 BY SIMILARITY

Query Match 42.7%; Score 41; DB 1; Length 432;  
 Best Local Similarity 70.0%; Pred. No. 21;  
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 RLPHFTGCG 16  
 Db 79 RCPHGVCGG 88

RESULT 13

GAG\_SRV1 STANDARD; PRT; 658 AA.  
 ID GAG\_SRV1  
 AC P04022;  
 DT 23-OCT-1986 (Rel. 02, Created)  
 DT 23-OCT-1986 (Rel. 02, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE GAG polypeptide (Core polypeptide) [Contains: Core protein P10; Core phosphoprotein P18; Core protein P12; Core protein P27; Core protein P14; Core protein P4].  
 GN GAG.  
 OS Simian retrovirus SRV-1.  
 OC Viruses; Retroviral viruses; Retroviridae.  
 OX NCBI\_TaxID=11942;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86151668; PubMed=1006247;  
 RA Power M.D., Marx P.A., Bryant M.L., Gardner M.B., Barr P.J.,  
 RA Luciw P.A.;  
 RT "Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus.";  
 RL Science 231:1567-1572(1986).  
 CC -1- SIMILARITY: CONTAINS 2 CCHC-TYPE ZINC FINGERS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M1841; AAA47730.1; -  
 DR PIR: A03951; FOLISA.  
 DR InterPro: IPR003322; Gag\_P10.  
 DR InterPro: IPR000721; Gag\_P24.  
 DR InterPro: IPR001878; ZnF\_CCHC.  
 DR Pfam: PF00098; zf-CCHC; 2.  
 DR Pfam: PF00607; Gag\_P24; 1.  
 DR Pfam: PF02337; Gag\_P10; 1.  
 DR PRINTS: PR00939; C2HCZNFINGER.  
 DR SMART: SM00343; ZnF\_C2HC; 2.  
 DR PROSITE: PS0158; ZF\_CCHC; 1.  
 KM Core protein; Polypeptide; Phosphorylation; Zinc-finger; Repeat.  
 FT CHAIN 1 162  
 FT CHAIN 163 217  
 FT CHAIN 218 300  
 FT CHAIN 301 526  
 FT CHAIN 527 622  
 FT CHAIN 623 658  
 FT CHAIN 658 658  
 FT ZN\_FING 548 565  
 CCHC-TYPE 1.

FT ZN\_FING 577 594 CCHC-TYPE 2.  
 SO SEQUENCE 658 AA; 73195 MW; 60929C787AF6923A CRC64;  
 Query Match 42.7%; Score 41; DB 1; Length 658;  
 Best Local Similarity 70.0%; Pred. No. 32;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GFSRLPHFTG 13  
 Db 268 GFSRLPHWP 277

RESULT 14

MML1\_MYCTU STANDARD; PRT; 958 AA.  
 ID MML1\_MYCTU  
 AC P95211;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative membrane protein mmpL1.  
 GN MmpL1 OR RV0402C OR MT0412 OR MTCY04D9.15C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv.  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J.F., Deboy R., Dodson R., Gwin M.L., Haft D., Hickey E.,  
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -----  
 CC -1- SIMILARITY: BELONGS TO THE MmpL FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z84725; CAB06593.1; -  
 DR EMBL: AE006945; AAK44636.1; -  
 DR TIGR: MT0412; -  
 DR TubercuList; RV0402C; -  
 DR InterPro: IPR004707; ActII.  
 DR InterPro: IPR004869; MmpL.  
 DR Pfam: PF03176; MmpL; 2.  
 DR TIGRFAMs: TIGR00833; actII; 1.  
 KM Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 19 39  
 FT TRANSMEM 192 212  
 POTENTIAL.

```
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 295 315 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 377 397 POTENTIAL.
FT TRANSMEM 762 782 POTENTIAL.
FT TRANSMEM 791 811 POTENTIAL.
FT TRANSMEM 814 834 POTENTIAL.
FT TRANSMEM 868 888 POTENTIAL.
FT TRANSMEM 890 910 POTENTIAL.
SQ SEQUENCE 958 AA; 104838 MW; 74FC0C8F54C3A14 CRC64;

Query Match
Best local Similarity 42.7%; Score 41; DB 1; Length 958;
Pred. No. 46;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LGFSRLPHF 11
1111111111
DB 313 LGFSRLPHY 321
```

## RESULT 15

YPJA\_ECOLI

```
ID YPJA_ECOLI STANDARD; PRT; 1569 AA.
AC P52143; P76610; P77017; P77019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane protein ypjA.
GN YPJA OR B2647.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
```

SEQUENCE FROM N.A.

STRAIN=K12 / MG1655;

MEDLINE=97426617; PubMed=9278503;

Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

Mau B., Shao Y.;

"The complete genome sequence of Escherichia coli K-12.";

Science 277:1453-1474(1997).

[2]

SEQUENCE FROM N.A.

STRAIN=K12;

MEDLINE=97349980; PubMed=9205837;

Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,

Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,

Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,

Oshima T., Oyama S., Saito N., Sampei G., Satch Y., Sivasundaram S.,

Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,

Yamagata S., Horiuchi T.;

"Construction of a contiguous 874-kb sequence of the Escherichia coli

K-12 genome corresponding to 50,0-68.8 min on the linkage map and

analysis of its sequence features.";

DNA Res. 4:91-113(1997).

-1- SUBCELLULAR LOCATION: Outer membrane (Potential).

-1- SIMILARITY: STRONG, TO BORDETELLA PERTACTIN.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL: U36840; AAA79815.1; ALT\_SEQ.

EMBL: AE000350; AAC75695.1; -

EMBL: D90889; BAA16514.1; ALT\_INIT.

EMBL: D90890; BAA16518.1; ALT\_INIT.

```
DR EcoGene; EG13213; ypjA.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF03212; Pertactin; 1.
KW Hypothetical protein; Outer membrane; Complete proteome.
SQ SEQUENCE 1569 AA; 162774 MW; 0D006A9A32154596 CRC64;
```

```
Query Match
Best local Similarity 42.7%; Score 41; DB 1; Length 1569;
Pred. No. 76;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 4 GFSRLPHFTGCGG 16
1111111111
DB 508 GYNKLSHFTTGG 520
```

Search completed: May 20, 2003, 16:35:37  
Job time : 6.57692 secs

**THIS PAGE BLANK (uspto)**



GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:19:09 ; Search time: 21.0165 seconds  
(without alignments)  
166.669 Million cell updates/sec

Title: US-09-869-155-14  
Perfect score: 96  
Sequence: 1 KRIGFSRLPFTGCGSL 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	61.5	419	10	08S1V0
2	48	50.0	187	2	08RKC9
3	46.5	48.4	198	10	094K29
4	46.5	48.4	253	10	023605
5	46	47.9	505	3	08X0V8
6	46	47.9	602	13	0902B0
7	45	46.9	630	4	096AD9
8	44	45.8	361	11	09CUE9
9	44	45.8	391	10	09CUE9
10	44	45.8	719	5	09N8Z8
11	43.5	45.3	438	10	09S756
12	43	44.8	206	16	08ZPM4
13	43	44.8	378	10	0943A3
14	43	44.8	410	6	097645
15	43	44.8	484	2	093548
16	42	43.8	203	10	09C7D9

17	42	43.8	255	16	068877	068877 pseudomonas
18	42	43.8	268	4	09NSS9	09NSS9 homo sapien
19	42	43.8	324	10	09M7E7	09M7E7 arabidopsis
20	42	43.8	325	10	080682	080682 arabidopsis
21	42	43.8	325	10	09LH52	09LH52 arabidopsis
22	42	43.8	325	10	0938C1	0938C1 arabidopsis
23	42	43.8	371	16	08Y1X4	08Y1X4 ralsstonia s
24	42	43.8	440	16	092LX3	092LX3 rhizobium m
25	42	43.8	472	10	09LZT1	09LZT1 arabidopsis
26	42	43.8	533	16	09KXZ1	09KXZ1 streptomyc
27	42	43.8	1444	5	093XW0	093XW0 caenorhabdi
28	42	43.8	1512	10	09M0S8	09M0S8 arabidopsis
29	42	43.8	3097	5	061143	061143 plasmodium
30	41.5	43.2	581	10	09SVP9	09SVP9 arabidopsis
31	41	42.7	96	12	084679	084679 parametium
32	41	42.7	178	3	014107	014107 schizosacch
33	41	42.7	211	16	09E0Q9	09E0Q9 ureaplasma
34	41	42.7	264	5	0905Z7	0905Z7 anophelies s
35	41	42.7	309	5	095XV7	095XV7 caenorhabdi
36	41	42.7	333	16	P73516	P73516 synecocyst
37	41	42.7	338	17	08TH25	08TH25 pyrococcus
38	41	42.7	374	10	0942D5	0942D5 oryza sativ
39	41	42.7	436	5	09N9R6	09N9R6 leishmania
40	41	42.7	457	5	017032	017032 caenorhabdi
41	41	42.7	465	5	09Y0Z9	09Y0Z9 drosophila
42	41	42.7	534	5	P92031	P92031 drosophila
43	41	42.7	534	5	09VJ37	09VJ37 drosophila
44	41	42.7	591	16	08YDM6	08YDM6 brucella me
45	41	42.7	595	10	094AM9	094AM9 arabidopsis

## ALIGNMENTS

RESULT 1						
Q8S1V0	PRELIMINARY;	PRT;	419	AA.		
AC 08S1V0:						
DT 01-JUN-2002 (TREMREL_21, Created)						
DT 01-JUN-2002 (TREMREL_21, Last sequence update)						
DT 01-JUN-2002 (TREMREL_21, Last annotation update)						
DE Putative dermal glycoprotein.						
GN P0504E02.9.						
OS Oryza sativa (japonica cultivar-group).						
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;						
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;						
OC Eriarthoideae; Oryzeae; Oryza.						
OX NCBI_TaxID=39947;						
RN [1]						
RP SEQUENCE FROM N.A.						
RC STRAIN=CV. NIPPONBARE;						
RA Sasaki T., Matsumoto T., Yamamoto K.;						
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC						
RT clone:P0504E02.9";						
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.						
DR EMBL: AP003269; BAB89708.1; -						
SO SEQUENCE	419	AA;	43786	MM;	792B5F5A4F3A8	CRC64;
Query Match	61.5%	Score	59;	DB	10;	Length 419;
Best Local Similarity	71.4%	Pred. No.	0.13;			
Matches	10;	Conservative	2;	Mismatches	2;	Indels 0;
						Gaps 0;
QY 1 KRIGFSRLPFTGCG 14						
DB 397 KRIGFSRLPFTGCG 410						
RESULT 2						
Q8RKC9	PRELIMINARY;	PRT;	187	AA.		
ID 08RKC9:						
DT 01-JUN-2002 (TREMREL_21, Created)						
DT 01-JUN-2002 (TREMREL_21, Last sequence update)						

```
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE VME protein.
GN VME.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC SRRATN-3937;
RA Reverchon S.;
RT "vfm genes of Erwinia chrysanthemi modulate the synthesis of multiple
RT virulence factors."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ438188; CAD27336.1;
SO SEQUENCE 187 AA; 21903 MW; 9F85CF180AC8EAD6 CRC64;

Query Match 50.0%; Score 48; DB 2; Length 187;
Best Local Similarity 72.7%; Pred. No. 4;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 4 GFSRLPHTCG 14
DB 119 GFSRLPHSRC 129

RESULT 3
O94K29 PRELIMINARY; PRT; 198 AA.
AC O94K29;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Putative sperm protein homolog.
GN DL4860W.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Iiu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C.C., Toriumi M., Yu G.,
RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shim P., Southwick A.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene dl4860w."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinozaki K., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RT "Full length cDNA of gene dl4860w."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF370345; AAK4160.1;
DR EMBL: AY062975; AAL34149.1;
DR InterPro: IPR005340; UPF0083.
DR Pfam: PF03654; UPF0083.1.
SO SEQUENCE 198 AA; 22694 MW; 763B9B8721884E40 CRC64;

Query Match 48.4%; Score 46.5; DB 10; Length 198;
Best Local Similarity 43.5%; Pred. No. 7.4;
Matches 10; Conservative 1; Mismatches 3; Indels 9; Gaps 1;
```

```
OY 2 RLGSRLP-----HFTGCG 15
DB 127 RVSEFRLPGSVSLQRRHFLCG 149

RESULT 4
ID 023605 PRELIMINARY; PRT; 253 AA.
AC 023605;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, last annotation update)
DE DNA chromosome 4, ESSA I CONTFG fragment NO. 8 (Sperm protein
DE homolog).
GN ATG417650.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98121113; Pubmed=9461215;
RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA Bergkamp R., Dirkse W., van Staveren M., Slekema W., Drost L.,
RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Wedler E., Wambutt R., Woltzenegger T., Pohl T.M., Terry N.,
RA Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA Audorg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
RA Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA Vouklatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA Halbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA Palmer K., Benes V., Reichman S., Ansgore W., Cooke R., Berger C.,
RA Delsen M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA Schueller C., Chalwatzis N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana."
RL Nature 391:485-488(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: 297343; CAB10545.1;
DR EMBL: AL161546; CAB8768.1;
SO SEQUENCE 253 AA; 28598 MW; D76F06B291657B9C CRC64;

Query Match 48.4%; Score 46.5; DB 10; Length 253;
Best Local Similarity 43.5%; Pred. No. 9.5;
Matches 10; Conservative 1; Mismatches 3; Indels 9; Gaps 1;

OY 2 RLGSRLP-----HFTGCG 15
DB 127 RVSEFRLPGSVSLQRRHFLCG 149

RESULT 5
ID 08X0V8 PRELIMINARY; PRT; 505 AA.
AC 08X0V8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE Hypothetical 55.6 kDa protein.
GN 123A4.340.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hohenseil J., Brandt P., Fartmann B., Holland R.,
RA Nykatura G., Mewes H.W., Mannhaupt G.;
```

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA German Neurospora genome project;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL670009; CAD21385.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 505 AA; 55613 MW; 4E88DDBA42383E7 CRC64;

Query Match 47.9%; Score 46; DB 3; Length 505;  
 Best Local Similarity 62.5%; Pred. No. 23;  
 Matches 10; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

OY 3 LGFSRLPHFTGC--CG 16  
 DB 229 LGFSRLPHFTGCAPDGC 244

## RESULT 6

O90ZB0 PRELIMINARY; PRT; 602 AA.  
 AC O90ZB0:  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Growth hormone receptor.  
 GN GHR.

OS Carassius auratus (Goldfish).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Carassius.  
 OX NCBI\_TaxID=7957;

RA MEDLINE-21297186; PubMed-11404009;  
 RA Lee L.T.O., Nong G., Chan Y.H., Tse D.L.Y., Cheng C.H.K.;  
 RA "Molecular cloning of a teleost growth hormone receptor and its  
 RT functional interaction with human growth hormone."  
 RL Gene 270:121-129(2001).

DR EMBL: AF293417; AAK60495.1; -  
 DR InterPro: IPR002996; CRIA.  
 DR InterPro: IPR003961; FN\_LIII.  
 DR InterPro: IPR003528; Hemtopoptn\_L\_F1.  
 DR SMART; SMO0060; FN3; 1.  
 DR PROSITE; PS01352; HEMATOPO\_REC\_L\_F1; UNKNOWN\_1.  
 KW Receptor.

SQ SEQUENCE 602 AA; 67262 MW; 4D9F66821C1A029E CRC64;

## Query Match

Best Local Similarity 47.9%; Score 46; DB 13; Length 602;  
 Pred. No. 27;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 6 SRLPHFTGC 14  
 DB 33 SRLPHFTGC 41

## RESULT 7

O96AD9 PRELIMINARY; PRT; 630 AA.  
 AC O96AD9:  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Hypothetical 66.8 kDa protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MUSCLE;  
 RA Strausberg R.;

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC017261; AAH17261.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 630 AA; 66805 MW; FFDD6F5F020FBC80 CRC64;

Query Match 46.9%; Score 45; DB 4; Length 630;  
 Best Local Similarity 77.8%; Pred. No. 41;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 8 LPHFTGCG 16  
 DB 135 LPHFTGCG 143

## RESULT 8

O9CUE9 PRELIMINARY; PRT; 361 AA.  
 AC O9CUE9:  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE 4930597B14Rik protein (Fragment).  
 DE 4930597B14Rik.  
 GN 4930597B14Rik.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
 RX MEDLINE-21085660; PubMed-11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kono S., Yamana K.,  
 RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasikawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuenli P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
 RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Mordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,  
 RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK016403; BAB30217.1; -  
 DR HSSP; P11940; ICVU.  
 DR MGD; MGI:1922638; 4930597B14Rik.  
 DR InterPro: IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 1.  
 DR SMART; SMO0360; RRM; 1.  
 DR PROSITE; PS0102; RRM; 1.  
 DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT 361  
 SQ SEQUENCE 361 AA; 39350 MW; DE4625D464C6FE30 CRC64;

Query Match 45.8%; Score 44; DB 11; Length 361;  
 Best Local Similarity 64.3%; Pred. No. 35;  
 Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 LGFSRLPHFTGCG 16  
 DB 18 LGFSRLPHFTGCG 31

## RESULT 9

O9CUE9 PRELIMINARY; PRT; 361 AA.  
 AC O9CUE9:  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE 4930597B14Rik protein (Fragment).  
 DE 4930597B14Rik.  
 GN 4930597B14Rik.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;

Q9FW53 PRELIMINARY; PRT; 391 AA.  
 ID Q9FW53  
 AC Q9FW53  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Putative protoporphyryinogen oxidase.  
 GN OSUNB0094K03.11.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eurythidaceae; Oryzaceae; Oryza.  
 OC NCB1\_TaxID=4530;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,  
 RA Hsiao J., Zismann V., Pal G., Bowman C.L., Fujii C.Y., Vanaken S.E.,  
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,  
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;  
 RT "Oryza sativa chromosome 10 BAC OSUNB0094K03 genomic sequence."  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC069145; AAC16857.1;  
 SO SEQUENCE 391 AA; 41019 MW; 7B7FA25F21931B39 CRC64;

Query Match 45.8%; Score 44; DB 10; Length 391;  
 Best Local Similarity 56.2%; Pred. No. 38;  
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 RLGFSRLPHTGCGCL 17  
 DB 358 RRGRRPRRPGCGGCL 373

RESULT 10  
 O9NR28 PRELIMINARY; PRT; 719 AA.  
 ID O9NR28  
 AC O9NR28;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical 79.9 kDa protein.  
 GN CH1.79.  
 OS Trypanosoma brucei.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCB1\_TaxID=5691;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=TRE0927;  
 RA Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,  
 RA Leonard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,  
 RA Gerard C., Rajandream M.A., Barrell B.G.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: A1355782; CAB95385.1;  
 DR InterPro: IPR000345; CytC\_heme\_bind.  
 DR InterPro: IPR000433; Znf\_ZZ.  
 DR SMART: SM00291; Znf\_ZZ; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN\_1.  
 DR PROSITE: PS01357; ZF\_ZZ\_1; UNKNOWN\_1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 719 AA; 79901 MW; D1EDD08BD2587AED CRC64;

Query Match 45.8%; Score 44; DB 5; Length 719;  
 Best Local Similarity 88.9%; Pred. No. 69;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 LGSFRLPHF 11  
 DB 352 LGSFRLPYF 360

RESULT 11  
 O9ST55

ID O9ST56 PRELIMINARY; PRT; 438 AA.  
 AC O9ST56;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Alpha galactosyltransferase (Fragment).  
 GN GALTRAN.  
 OS Trigonella foenum-graecum.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Trigonella.  
 OX NCB1\_TaxID=78534;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC MEDLINE=20040039; PubMed=10571854;  
 RA Edwards M.E., Dickson C.A., Chengappa S., Sidebottom C.M.,  
 RA Gidley M.J., Reid G.;  
 RT "Molecular Characterisation of a membrane-bound galactosyltransferase  
 of plant cell wall matrix biosynthesis."  
 RL Plant J. 19:691-697(1999).  
 DR EMBL: AJ245478; CAB52246.1;  
 KW Glycosyltransferase; Transferase.  
 FT NON\_TER  
 SO SEQUENCE 438 AA; 51282 MW; E6E35B71AD21848A CRC64;

Query Match 45.3%; Score 43; DB 10; Length 438;  
 Best Local Similarity 58.8%; Pred. No. 51;  
 Matches 10; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

OY 1 KRLGFSRLP--HPTGC 14  
 DB 363 KNLGDMRRPFTHTGCG 379

RESULT 12  
 O8ZPM4 PRELIMINARY; PRT; 206 AA.  
 ID O8ZPM4  
 AC O8ZPM4;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Putative Na<sup>+</sup>-transporting NADH:ubiquinone oxidoreductase gamma  
 DE subunit.  
 GN YDGP OR STM1455.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCB1\_TaxID=602;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=LT72 / SGCJ412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Portwill S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grevail N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 LT2."  
 RL Nature 413:852-856(2001).  
 DR EMBL: AE008763; AAL20377.1;  
 KW Ubiquinone; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 206 AA; 22238 MW; 5E6BB49636F05C54 CRC64;

Query Match 44.8%; Score 43; DB 16; Length 206;  
 Best Local Similarity 42.9%; Pred. No. 29;  
 Matches 9; Conservative 2; Mismatches 4; Indels 6; Gaps 1;

OY 1 KRLGFSRLP-----SRLPHTGCG 15  
 DB 185 KRAGLYAESLPADLPHTACG 205

Query Match	44.8%	Score 43	DB 10	Length 378
Best Local Similarity	88.9%	Pred. No. 54		
Matches	8	Conservative	0	Mismatches 1
				Indels 0
				Gaps 0
OY	5	FSRLPFTG	13	
Db	119	FSRLPFTG	127	

RESULT 14			
097645			
ID	097645	PRELIMINARY:	PRT: 410 AA.
AC	097645;		
DT	01-MAV-1999 (TREMBLrel. 10, Created)		
DT	01-MAV-1999 (TREMBLrel. 10, last sequence update)		
DT	01-NOV-1999 (TREMBLrel. 12, last annotation update)		
DE	Fibrinogen A-alpha chain (Fragment).		
OS	Macropus rufus (Red kangaroo) (Megalalia rufa).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi		
OC	Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.		
OX	NCBI_TaxID=9321;		
RN	[1]		
RP			
RA	SEQUENCE FROM N.A.		
RL	Murakawa M.;		
DR	Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: AF095467; AAC67566.1; -		
DR	HSSP: P02671; 1FZF.		
FT	NON_TER	1	1
FT		410	410
SQ	SEQUENCE	410 AA; 43891 MW; 7C81EAD5FC849E8 CRC64;	

Query Match	44.8%	Score 43	DB 6	Length 410
Best Local Similarity	64.7%	Pred. No. 58		
Matches 11	Conservative 0	Mismatches 4	Indels 2	Gaps 1
OY	3	LGFSRLPH-FTGCCGL	17	
db	294	LSFLRLTHYFFFGRCGL	310	

RESULT 15		
093548		
ID 093548	PRELIMINARY;	PRT; 484 AA.
AC 093548;		
DT 01-DEC-2001	(TREMBLrel, 19, Created)	
DT 01-DEC-2001	(TREMBLrel, 19, Last sequence update)	

```

DT      01-Dec-2001 (TReMBLrel_19, Last annotation update)
DE      Putative membrane protein.
OS      Myxococcus xanthus.
OC      Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC      Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
OX      NCBI_TaxID=34;
RP      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=89123159; PubMed=2464581;
RA      Stephens K., Hartzell P., Kaiser D.;
RT      "Gliding motility in Myxococcus xanthus: mgl locus, RNA, and predicted
RT      protein products."
RL      J. Bacteriol. 171:819-830(1989).
RN      [2]
RN      SEQUENCE FROM N.A.
RP      Thomasson B., Stassinopoulos A., White D.J., Plamann L.,
RA      Hartzell P.L.;
RT      "The small GTPase, MglA, interacts with a protein kinase to control
RT      social motility."
RL      Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AF377950; AAK54651.1; -.
SO      SEQUENCE 484 AA: 50751 MW: 108A19C9C9CC09DB CXC64;

Query Match          44.8%; Score 43; DB 2; Length 484;
Best Local Similarity 77.8%; Pred. No. 68;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY      4 GFSRUPHT 12
db      111:111:
100 GFSOLPHES 108

```

Search completed: May 20, 2003, 16:43:10  
Job time : 24.0165 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:35:02 ; Search time 7.84615 seconds  
(without alignments)  
63.750 Million cell updates/sec

Title: US-09-869-155-14

Perfect score: 96

Sequence: 1 KRIGFSRLPHFTGCGCL 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	45.8	932	3	US-08-968-752B-6
2	44	45.8	932	4	US-09-536-224-6
3	44	44.8	163	4	US-09-562-737-76
4	42	43.8	163	4	US-09-562-737-71
5	41	42.7	163	4	US-09-562-737-72
6	41	42.7	163	4	US-09-562-737-75
7	41	42.7	729	4	US-09-291-922-28
8	40	41.7	163	4	US-09-562-737-79
9	40	41.7	902	4	US-09-644-600-10
10	39	40.6	64	4	US-09-288-143-155
11	38	39.6	40	4	US-09-046-894-39
12	38	39.6	76	1	US-08-702-344-6
13	38	39.6	335	4	US-09-797-906-2
14	38	39.6	336	4	US-09-141-206-7
15	38	39.6	499	2	US-08-820-170A-40
16	38	39.6	499	3	US-09-055-699-40
17	38	39.6	499	4	US-09-273-565-40
18	38	39.6	499	4	US-09-565-538-40
19	38	39.6	499	4	US-09-661-468-40
20	38	39.6	537	4	US-09-138-277C-3
21	38	39.6	558	4	US-09-138-277C-1
22	38	39.6	699	4	US-08-274-121B-6
23	38	39.6	737	4	US-09-291-922-8
24	38	39.6	855	2	US-09-027-337-2
25	38	39.6	855	4	US-09-644-600-2
26	38	39.6	922	4	US-09-141-206-6
27	38	39.6	933	4	US-09-141-206-2

28	38	39.6	933	4	US-09-107-149-17	Sequence 17, Appl
29	38	39.6	1841	2	US-08-804-227C-6	Sequence 6, Appl
30	38	39.6	5215	4	US-09-105-537-2	Sequence 2, Appl
31	37.5	39.1	414	4	US-09-134-001C-5528	Sequence 5528, Ap
32	37	38.5	2071	4	US-09-415-522-6	Sequence 6, Appl
33	36.5	38.0	1151	4	US-09-177-165A-31	Sequence 31, Appl
34	36	37.5	31	3	US-09-100-414B-89	Sequence 89, Appl
35	36	37.5	31	4	US-09-303-323-69	Sequence 80, Appl
36	36	37.5	163	4	US-09-562-737-78	Sequence 78, Appl
37	36	37.5	163	4	US-09-562-737-80	Sequence 18, Appl
38	36	37.5	237	1	US-08-750-532-18	Sequence 3456, Ap
39	36	37.5	276	4	US-09-134-001C-3456	Sequence 1, Appl
40	36	37.5	392	2	US-08-886-152-1	Sequence 3, Appl
41	36	37.5	392	4	US-08-886-152-3	Sequence 1, Appl
42	36	37.5	392	4	US-09-196-232-1	Sequence 3, Appl
43	36	37.5	392	4	US-09-196-232-3	Sequence 31, Appl
44	36	37.5	579	1	US-08-126-564A-31	
45	36	37.5	579	5	PCT-US94-09143-31	

ALIGNMENTS

```
RESULT 1
US-08-968-752B-6
: Sequence 6, Application US/08968752B
: Patent No. 6043073
:
: GENERAL INFORMATION:
: APPLICANT: Frohman, Michael A.
: APPLICANT: Morris, Andrew
: TITLE OF INVENTION: No. 6043073el Phospholipase D Polypeptide and
: TITLE OF INVENTION: DNA Sequences
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ONYX Pharmaceuticals, Inc.
: STREET: 3031 Research Drive
: CITY: Richmond
: STATE: California
: COUNTRY: USA
: ZIP: 94806
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/968,752B
: FILING DATE: 13-AUG-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/025,469
: FILING DATE: 05-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Giotta, Gregory J.
: REGISTRATION NUMBER: 32,028
: REFERENCE/DOCKET NUMBER: ONYX2004
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 510-222-9700
: TELEFAX: 510-222-9758
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 932 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-968-752B-6
:
: Query Match 45.8%; Score 44; DB 3; Length 932;
: Best Local Similarity 60.0%; Pred. No. 37;
: Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
:
: 1 KRIGFSRLPHFTGCG 15
: || | | | | | | |
```

DE 211 KRSGHRVPGFTFCG 225

RESULT 2  
US-09-536-224-6

Sequence 6, Application US/09536224  
Patent No. 6379665  
GENERAL INFORMATION:  
APPLICANT: Frohman, Michael A.  
TITLE OF INVENTION: No. 6379665el Phospholipase D polypeptide and  
DNA Sequences  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ONYX Pharmaceuticals, Inc.  
STREET: 3031 Research Drive  
CITY: Richmond  
STATE: California  
COUNTRY: USA  
ZIP: 94806  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/536,224  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/968,752  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Giotta, Gregory J.  
REGISTRATION NUMBER: 32,028  
REFERENCE/DOCKET NUMBER: ONYX2004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 510-222-9700  
TELEFAX: 510-222-9758  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 932 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-536-224-6

Query Match 45.8%; Score 44; DB 4; Length 932;  
Best Local Similarity 60.0%; Pred. No. 37;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 KRGSRLPHFTCG 15  
DB 211 KRSGHRVPGFTFCG 225RESULT 3  
US-09-562-737-76

Sequence 76, Application US/09562737  
Patent No. 6428967  
GENERAL INFORMATION:  
APPLICANT: Herz, Joachim  
APPLICANT: Gotthardt, Michael  
TITLE OF INVENTION: LDL Receptor Signaling Pathways  
FILE REFERENCE: UTSW0708  
CURRENT APPLICATION NUMBER: US/09/562,737  
CURRENT FILING DATE: 2000-05-01  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 76  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Artificial Sequence

FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
OTHER INFORMATION: Sequence  
US-09-562-737-76

Query Match 44.8%; Score 43; DB 4; Length 163;  
Best Local Similarity 66.7%; Pred. No. 8.5;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KRGSRLPHFT 12  
DB 84 KMGFERLEHFS 95

## RESULT 4

US-09-562-737-71  
Sequence 71, Application US/09562737  
Patent No. 6428967  
GENERAL INFORMATION:  
APPLICANT: Herz, Joachim  
APPLICANT: Gotthardt, Michael  
TITLE OF INVENTION: LDL Receptor Signaling Pathways  
FILE REFERENCE: UTSW0708  
CURRENT APPLICATION NUMBER: US/09/562,737  
CURRENT FILING DATE: 2000-05-01  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 71  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-562-737-71

Query Match 43.8%; Score 42; DB 4; Length 163;  
Best Local Similarity 61.5%; Pred. No. 12;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KRGSRLPHFTG 13  
DB 84 KMGFERLEHFSG 96

## RESULT 5

US-09-562-737-72  
Sequence 72, Application US/09562737  
Patent No. 6428967  
GENERAL INFORMATION:  
APPLICANT: Herz, Joachim  
APPLICANT: Gotthardt, Michael  
TITLE OF INVENTION: LDL Receptor Signaling Pathways  
FILE REFERENCE: UTSW0708  
CURRENT APPLICATION NUMBER: US/09/562,737  
CURRENT FILING DATE: 2000-05-01  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 72  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-562-737-72

Query Match 42.7%; Score 41; DB 4; Length 163;  
Best Local Similarity 61.5%; Pred. No. 18;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KRGSRLPHFTG 13  
DB 84 KMGFERLEHFSG 96



```
Db      84  KMGFGRLNHFSG 96

RESULT 6
US-09-562-737-75
; Sequence 75, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 75
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-75

Query Match      42.7%; Score 41; DB 4; Length 163;
Best Local Similarity 61.5%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY      1  KRLGFSRLPHFTG 13
      1  |||||
Db      84  KMGFGRLNHFSG 96

RESULT 7
US-09-291-922-29
; Sequence 29, Application US/09291922
; Patent No. 6383776
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Hiltz, Bill
; APPLICANT: Kinney, Tony
; APPLICANT: Tingey, Scott
; TITLE OF INVENTION: Plant Sugar Transport Proteins
; FILE REFERENCE: BB-1163
; CURRENT APPLICATION NUMBER: US/09/291,922
; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 60/083,044
; EARLIER FILING DATE: April 24, 1998
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 29
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-291-922-29

Query Match      42.7%; Score 41; DB 4; Length 729;
Best Local Similarity 70.0%; Pred. No. 88;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY      8  LPHFTGCGGL 17
      1  |||||
Db      136  LPHFTGSGGM 145

RESULT 8
US-09-562-737-79
; Sequence 79, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael

; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-79

Query Match      41.7%; Score 40; DB 4; Length 163;
Best Local Similarity 61.5%; Pred. No. 26;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY      1  KRLGFSRLPHFTG 13
      1  |||||
Db      84  KMGFGRLNHFSG 96

RESULT 9
US-09-644-600-10
; Sequence 10, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 10
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: Epithin
US-09-644-600-10

Query Match      41.7%; Score 40; DB 4; Length 902;
Best Local Similarity 38.1%; Pred. No. 1.6e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 4; Gaps 1;

OY      1  KRLGFSRLPHFTG 17
      1  |||||
Db      323  RHLGFEATFQPLPKMSSCGV 343

RESULT 10
US-09-288-143-155
; Sequence 155, Application US/09288143
; Patent No. 6433139
; GENERAL INFORMATION:
; APPLICANT: Brewer et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P1
; CURRENT APPLICATION NUMBER: US/09/288,143
; CURRENT FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/US98/21142
; EARLIER FILING DATE: 1998-10-08
; EARLIER APPLICATION NUMBER: 60/061,463
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,529
```

EARLIER FILING DATE: 1997-10-09  
EARLIER APPLICATION NUMBER: 60/071,498  
EARLIER FILING DATE: 1997-10-09  
EARLIER APPLICATION NUMBER: 60/061,527  
EARLIER FILING DATE: 1997-10-09  
EARLIER APPLICATION NUMBER: 60/061,536  
EARLIER FILING DATE: 1997-10-09  
EARLIER APPLICATION NUMBER: 60/061,532  
EARLIER FILING DATE: 1997-10-09  
NUMBER OF SEQ ID NOS: 219  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 155  
LENGTH: 64  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-288-143-155

Query Match 40.6%; Score 39; DB 4; Length 64;  
Best Local Similarity 47.6%; Pred. No. 14;  
Matches 10; Conservative 2; Mismatches 5; Indels 4; Gaps 1;

OY 1 KRIGSRLPHT---GCGGL 17  
1 : 111 11:1 111  
16 KWIGSFIFPHWTWIDLEIGL 36

Dr 16 KWIGSFIFPHWTWIDLEIGL 36

RESULT 11  
US-09-046-894-39  
Sequence 39, Application US/09046894  
Patent No. 6190857  
GENERAL INFORMATION:  
APPLICANT: Ralph, David  
APPLICANT: An, Gang  
APPLICANT: O'Hara, Mark S.  
APPLICANT: Veltl, Robert  
TITLE OF INVENTION: DIAGNOSIS OF DISEASE STATE USING mRNA  
TITLE OF INVENTION: PROFILES IN PERIPHERAL LEUKOCYTES  
NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/046,894  
FILING DATE: Concurrently Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/041,576  
FILING DATE: 24-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Nakashima, Richard A.  
REGISTRATION NUMBER: P-42,023  
REFERENCE/DOCKET NUMBER: UROC:014  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 40 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-046-894-39

Query Match 39.6%; Score 38; DB 4; Length 40;

Best Local Similarity 85.7%; Pred. No. 12;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 PHETGGC 15  
11 11111  
DB 11 PHOTGGC 17

RESULT 12  
US-08-702-344-6  
Sequence 6, Application US/08702344  
Patent No. 5723315  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John  
APPLICANT: Lavallee, Edward  
APPLICANT: Racie, Lisa  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/702,344  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-702-344-6

Query Match 39.6%; Score 38; DB 1; Length 76;  
Best Local Similarity 70.0%; Pred. No. 25;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 4 GFSRLPHFTG 13  
11111111  
DB 10 GFSYLNHFSG 19

RESULT 13  
US-09-797-906-2  
Sequence 2, Application US/09797906  
Patent No. 6329188  
GENERAL INFORMATION:  
APPLICANT: Zianghe YAN, Karen A. KETCHUM, Valentina DIPIRANCESCO, Ellen M. BEASLEY  
TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
FILE REFERENCE: C0001151CIP

;; CURRENT APPLICATION NUMBER: US/09/797,906  
;; CURRENT FILING DATE: 2001-03-05  
;; NUMBER OF SEQ ID NOS: 5  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO: 2  
;; LENGTH: 335  
;; TYPE: PRT  
;; ORGANISM: Human  
US-09-797-906-2

Query Match 39.6%; Score 38; DB 4; Length 335;  
Best Local Similarity 58.3%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 GFGRLPHTGCG 15  
DB 242 GFVCPHFVGHG 253

RESULT 14  
US-09-141-206-7  
; Sequence 7, Application US/09141206  
; Patent No. 6187559  
; GENERAL INFORMATION:  
; APPLICANT: Steed, Paul M.  
; APPLICANT: Lasala, Daniel J.  
; TITLE OF INVENTION: Amino Acid Sequence of Human PLD2A  
; Patent No. 6187559  
; FILE REFERENCE: 4-30148/PL/CGC1954/R  
; CURRENT APPLICATION NUMBER: US/09/141,206  
; CURRENT FILING DATE: 1998-08-27  
; EARLIER APPLICATION NUMBER: 60/057,802  
; EARLIER FILING DATE: 1997-08-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 7  
; LENGTH: 336  
; TYPE: PRT  
; ORGANISM: Human  
US-09-141-206-7

Query Match 39.6%; Score 38; DB 4; Length 336;  
Best Local Similarity 53.3%; Pred. No. 1.2e+02;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 KRGFGRLPHTGCG 15  
DB 211 KRSGHRVPELTCCG 225

RESULT 15  
US-08-820-170A-40  
; Sequence 40, Application US/08820170A  
; Patent No. 5831058  
; GENERAL INFORMATION:  
; APPLICANT: Tsutomu, FUJIWARA  
; APPLICANT: Takeshi, WATANABE  
; APPLICANT: Masato, HORIE  
; APPLICANT: Toyomasa, KATAGIRI  
; TITLE OF INVENTION: HUMAN GENE  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/820,170A  
;; FILING DATE:  
;; CLASSIFICATION: 536  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 293-7060  
;; TELEFAX: (202) 293-7860  
;; TELEX: 6491103  
;; INFORMATION FOR SEQ ID NO: 40:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 499 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-820-170A-40

Query Match 39.6%; Score 38; DB 2; Length 499;  
Best Local Similarity 47.4%; Pred. No. 1.8e+02;  
Matches 9; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 1 KRGFGR--RLPHFTGCGGL 17  
DB 246 KRVGFSPKKDIHPMPCSL 264

Search completed: May 20, 2003, 16:44:38  
Job time : 8.84615 secs

**THIS PAGE BLANK (USPTO)**



XX Claim 24; Page 112; 112pp; English.  
XX  
XX The present sequence is derived from an endo-beta-1,4-xyylanase  
CC inhibitor. The protein is obtained from wheat flour. The specification  
CC also describes a mutant xyylanase protein. The xyylanase is useful for  
CC preparing a foodstuff, preferably a bakery product or a substance  
CC (e.g. a dough) for making the bakery product. Wild type xyylanase or  
CC mutant xyylanase is useful for preparing a dough that is less sticky  
CC than a dough comprising a fungal xyylanase. The xyylanase inhibitor is  
CC useful for screening high degree resistance xyylanases for dough  
CC preparation. The xyylanase is also useful for preparing a non-sticky  
CC dough. A combination of xyylanase and the inhibitor is useful for  
CC calibrating and/or determining the quantity of inhibitor in a wheat  
CC flour sample.  
XX  
XX Sequence 21 AA:  
SQ  
Query Match 100.0%; Score 114; DB 21; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8,4e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LPVPAPVTKDPATSLYTIPEH 21  
Dh 1 LPVPAPVTKDPATSLYTIPEH 21  
RESULT 2  
ID AAE14689 standard; peptide: 21 AA.  
XX AAE14689;  
XX  
XX 21-AUG-2002 (first entry)  
XX  
XX Wheat flour xyylanase inhibitor B chain N-terminal fragment.  
DE  
XX Refrigerated dough; syruping; arabinoxylan; bakery product; bread;  
KW pizza base; cake; biscuit; wheat; flour; xyylanase inhibitor.  
XX  
XX Triticum aestivum.  
OS  
XX WO200152657-A1.  
PN  
XX 26-JUL-2001.  
PD  
XX 17-JAN-2001; 2001WO-IB00168.  
PF  
XX 18-JAN-2000; 2000GB-0001136.  
PR  
XX (DANI-) DANISCO AS.  
PA  
XX  
XX Poulsen CH, Sorensen JF;  
PI  
XX WPI: 2001-457446/49.  
DR  
XX  
XX Production of refrigerated dough with reduced syruping, useful in  
PT production of bakery products such as bread, comprises admixing cereal  
PT flour, water and protein that prevents enzymatic degradation of  
PT arabinoxylan in the cereal flour.  
PT  
XX Disclosure; Page 23; 26pp; English.  
PS  
XX The invention relates to a process for producing refrigerated dough  
CC with reduced 'syruping' (precipitation of liquid on the dough surface  
CC because of a reduction in water holding capacity caused by the breakdown  
CC of arabinoxylan over time). The process comprises admixing cereal flour  
CC and water with a protein that reduces/prevents enzymatic degradation of  
CC arabinoxylan in the cereal flour. The preferred protein is a xyylanase  
CC inhibitor. The method is useful to produce refrigerated dough in which  
CC syruping is reduced or eliminated. Refrigerated dough is typically  
CC stored for long periods to enable fresh baked products (e.g. bread,  
CC pizza bases, cakes, biscuits) to be produced as required. It may also be

CC possible to produce dough or baked products tailored to specific  
CC requirements using the method by the use of specific proteins/protein  
CC combinations. The present sequence is wheat flour  
CC endo-beta-1,4-xyylanase inhibitor B chain N-terminal fragment.  
XX  
XX Sequence 21 AA:  
SQ  
Query Match 100.0%; Score 114; DB 22; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8,4e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LPVPAPVTKDPATSLYTIPEH 21  
Dh 1 LPVPAPVTKDPATSLYTIPEH 21  
RESULT 3  
ID AAU07394 standard; protein: 21 AA.  
XX AAU07394;  
XX  
XX 18-DEC-2001 (first entry)  
XX  
XX Bacillus subtilis xyylanase inhibitor #3.  
DE  
XX Xylanase; plant cell wall; baking; cereal; starch production; wood;  
KW wood pulp bleaching; hemicellulose; arabinoxylan; foodstuff; inhibitor.  
XX  
XX Bacillus subtilis.  
OS  
XX WO200166711-A1.  
PN  
XX 13-SEP-2001.  
PD  
XX 08-MAR-2001; 2001WO-IB00426.  
PF  
XX 08-MAR-2000; 2000GB-0005585.  
PR  
XX 27-JUN-2000; 2000GB-0015751.  
PR  
XX (DANI-) DANISCO AS.  
PA  
XX  
XX Slbbsen O, Sorensen JF;  
PI  
XX WPI: 2001-596834/67.  
DR  
XX  
XX Novel variant xyylanase polypeptide or its fragment useful for degrading  
PT or modifying plant cell wall, comprises amino acid modifications such  
PT that the polypeptide has altered sensitivity to xyylanase inhibitor  
PT  
XX Disclosure; Page 63; 70pp; English.  
PS  
XX The invention relates to a variant xyylanase polypeptide (I) or its  
CC fragment having xyylanase activity, comprising one or more amino acid  
CC modifications such that (I) or its fragment has an altered sensitivity to  
CC a xyylanase inhibitor as compared with the parent xyylanase enzyme. (I) or  
CC its coding sequence (II) is useful for degrading or modifying plant cell  
CC wall or for processing a plant material by contacting the plant cell wall  
CC or plant material with (I) or (II). (I) is useful for modifying plant  
CC materials, and in baking, processing cereals, starch production,  
CC processing wood and enhancing the bleaching of wood pulp. (I)  
CC is useful for altering the viscosity derived from the presence of  
CC hemicellulose or arabinoxylan in a solution or system comprising plant  
CC cell wall material. (I) is useful for preparing a foodstuff such as  
CC bread, pretzels, tortillas, cakes, cookies, biscuits or crackers.  
CC The present sequence represents the amino acid sequence of Bacillus  
CC subtilis xyylanase inhibitor #3 as described in the method of the  
CC invention.  
XX  
XX Sequence 21 AA:  
SQ  
Query Match 100.0%; Score 114; DB 22; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8,4e-10;



22-JUN-2000; 2000GB-0015296.  
25-JAN-2001; 2001GB-0002018.  
26-JAN-2001; 2001GB-0002194.  
16-MAR-2001; 2001GB-0006564.  
21-MAY-2001; 2001GB-0012328.  
(LEUV-) LEUVEN RES & DEV.  
Delcours J, Debysse W, Gebuere K, Goesaert H, Fierens K, Robben J,  
Van Campenhout S;  
WPI: 2002-114579/15.  
Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
beta-glucanolytic enzymes comprises using endoxylanases during  
screening for inhibition activity or affinity chromatography with  
immobilised enzymes -  
Claim 127; Page 51; 127pp; English.  
The invention relates to separating and/or isolating inhibitors of  
cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
screening the inhibition activity by using two or more enzymes during the  
separation and/or isolation steps that allow to distinguish inhibitors of  
different specificity or by using an affinity chromatographic step with  
immobilised enzymes and/or antibodies against inhibitors. Also  
included are an isolated nucleic acid molecule encoding an inhibitor  
which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
nucleic acid, an expression system transformed with the nucleic  
acid, a host organism transformed with the nucleic acid, the inhibitory  
proteins encoded by the nucleic acids and modulators of the proteins.  
A recombinant protein, glycoprotein or polypeptide or microorganisms,  
plant or plant materials transformed with the nucleic acid are useful  
for the formation of an endoxylanase-inhibitor complex, screening  
endoxylanases that are totally, less or not inhibited by the inhibitors,  
reducing syringing in refrigerated dough compositions, affecting the  
relative affinity and/or relative hydrolysis specificity and/or relative  
hydrolysis rate versus water-extractable and/or water-unextractable  
arabinoxylans of endoxylanases such as by the formation of an  
endoxylanase/inhibitor complex, improving the malting of cereals such as  
barley, sorghum and wheat and/or the production of beer, improving the  
production and/or quality of baked or extruded cereal products such as  
straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
biscuits, pasta and noodles, animal feed stuff, improving the production  
of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
starch separation and production, improving maize processing,  
plant disease resistance and nutraceutical and/or pharmaceutical  
applications, improving paper and pulp technologies. The present  
sequence is a variant of wheat TXM1-I ('T. aestivum L. endoxylanase  
inhibitor').  
Note: Variant amino acids are highlighted in the specification but  
no wild-type sequence is shown for comparison.

DT	23-APR-2002	(first entry)
XX		
DE	Wheat L endoxylanase inhibitor, TAXI I, microheterogenic variant.	
XX		
KW	Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;	
KW	xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;	
KW	immobilised enzyme; enzyme; dough syruuping; cereal product; beer; plant;	
KW	straight dough; sponge dough; Chorleywood bread; biscuits; pasta;	
KW	noodle; animal feed; starch separation; maize processing; malting;	
KW	plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.	
XX		
OS	Triticum aestivum.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 95	/label= Unknown
FT	Misc-difference 98	/label= Unknown
FT	Misc-difference 101	/label= Unknown
FT	Misc-difference 110	/label= Unknown
FT	Misc-difference 145	/label= Unknown
FT	Misc-difference 183	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 232	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 275	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 282	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 333	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 333	/label= Unknown
XX		
PN	WO200198474-A1.	
XX		
DD	27-DEC-2001.	
XX		
PF	21-JUN-2001; 2001WO-BE00106.	
XX		
PR	22-JUN-2000; 2000GB-0015296.	
PR	25-JAN-2001; 2001GB-0002018.	
PR	26-JAN-2001; 2001GB-0002194.	
PR	16-MAR-2001; 2001GB-0006564.	
PR	21-MAY-2001; 2001GB-0012328.	
XX		
PA	(LEUV-) LEUVEN RES & DEV.	
XX		
PI	Delcours J, Debysse W, Gebruens K, Goesaert H, Flerens K, Robben J;	
PI	Van Campenhout S;	
XX		
DR	WPI; 2002-114579/15.	
XX		
PT	Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or	
PT	beta-glucanolytic enzymes comprising using endoxylanases during	
PT	screening for inhibition activity or affinity chromatography with	
PT	immobilised enzymes	
XX		
PS	Claim 127; Page 51; 127pp; English.	
XX		
CC	The invention relates to separating and/or isolating inhibitors of	
CC	cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising	
CC	screening the inhibition activity by using two or more enzymes during the	
CC	separation and/or isolation steps that allow to distinguish inhibitors of	
CC	different specificity or by using an affinity chromatographic step with	
CC	immobilised enzymes and/or antibodies against inhibitors. Also	
CC	included are an isolated nucleic acid molecule encoding an inhibitor	
CC	which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,	
CC	alpha-L-arabino-furanosidase and/or other cellulose, xylan,	
CC	arabinoxylan or beta-glucan degrading enzymes, a vector comprising the	
CC	nucleic acid, an expression system transformed with the nucleic	
CC	acid, a host organism transformed with the nucleic acid, the inhibitory	
CC		



CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a microheterogenic variant of wheat TAXI-I (T. aestivum L  
 CC endoxylanase inhibitor).  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.

XX Sequence 381 AA;  
 Query Match 91.2%; Score 104; DB 23; Length 381;  
 Best Local Similarity 95.2%; Pred. No. 5.6e-07;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 LPVPAPVTKDPATSLYTIPTF 21  
 DB 1 LPVLAPVTKDPATSLYTIPTF 21

RESULT 7  
 AAU75806 standard; peptide: 24 AA.

AC AAU75806;

DT 23-APR-2002 (first entry)

XX Wheat L endoxylanase inhibitor, TAXI I, N-terminus #1.

DE Wheat; TAXI-I; L endoxylanase inhibitor; cellulosytic enzyme inhibitor;

KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;

KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;

KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;

KW noodle; animal feed; starch separation; maize processing; malting;

KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

OS Triticum aestivum.

XX Key Location/Qualifiers

FT Misc-difference 21 /label= OTHER

FT /note= "Other= Any amino acid, preferably Leu"

FT Misc-difference 23 /label= OTHER

FT /note= "Other= Any amino acid, preferably Leu"

XX WO200198474-A1.

XX 27-DEC-2001.

XX 21-JUN-2001; 2001WO-BE00106.

XX 22-JUN-2000; 2000GB-0015296.

XX 25-JAN-2001; 2001GB-0002018.

XX 26-JAN-2001; 2001GB-0002194.

XX 16-MAR-2001; 2001GB-0006564.

XX 21-MAY-2001; 2001GB-001328.

PA (LEUV-) LEUVEN RES & DEV.

XX Delcourt J, Delyser W, Gebruers K, Goesaert H, Fierens K, Robben J;

PI Van Campenhout S;

XX WPI: 2002-114579/15.

DR Separating and/or isolating inhibitors of cellulosytic, xylanolytic, or

PT beta-glucanolytic enzymes comprises using endoxylanases during

PT screening for inhibitor activity or affinity chromatography with

PT immobilised enzymes

PS Claim 127; Page 9; 127pp; English.

XX The invention relates to separating and/or isolating inhibitors of

CC cellulosytic, xylanolytic and/or beta-glucanolytic enzymes comprises

CC screening the inhibition activity by using two or more enzymes during the

CC separation and/or isolation steps that allow to distinguish inhibitors of

CC different specificity or by using an affinity chromatographic step with

CC immobilised enzymes and/or antibodies against inhibitors. Also

CC included are an isolated nucleic acid molecule encoding an inhibitor

CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,

CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,

CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the

CC nucleic acid, an expression system transformed with the nucleic

CC acid, a host organism transformed with the nucleic acid, the inhibitory

CC proteins encoded by the nucleic acids and modulators of the proteins.

CC A recombinant protein, glycoprotein or polypeptide or microorganisms,

CC plant or plant materials transformed with the nucleic acid are useful

CC for the formation of an endoxylanase-inhibitor complex, screening

CC endoxylanases that are totally, less or not inhibited by the inhibitors,

CC reducing syruping in refrigerated dough compositions, affecting the

CC relative affinity and/or relative hydrolysis specificity and/or relative

CC hydrolysis rate versus water-extractable and/or water-unextractable

CC arabinoxylans of endoxylanases such as by the formation of an

CC endoxylanase/inhibitor complex, improving the malting of cereals such as

CC barley, sorghum and wheat and/or the production of beer, improving the

CC production and/or quality of baked or extruded cereal products such as

CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,

CC biscuits, pasta and noodles, animal feed stuff, improving the production

CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten

CC starch separation and production, improving maize processing,

CC plant disease resistance and nutraceutical and/or pharmaceutical

CC applications, improving paper and pulp technologies. The present

CC sequence is the N-terminus of wheat TAXI-I (T. aestivum L endoxylanase

CC inhibitor) from a 40-43kDa isolate of the protein.

XX Sequence 24 AA;

Query Match 84.2%; Score 96; DB 23; Length 24;

Best Local Similarity 95.0%; Pred. No. 3.9e-07;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 LPVPAPVTKDPATSLYTIPTF 20  
 DB 1 LPVLAPVTKDPATSLYTIPTF 20

RESULT 8  
 AAU75828 standard; Peptide: 22 AA.

AC AAU75828;

DT 23-APR-2002 (first entry)

XX Wheat xylanase inhibitor TAXI-I PCR product N-terminal sequence.

DE Wheat; TAXI-I; L endoxylanase inhibitor; cellulosytic enzyme inhibitor;

KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;

KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;

KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;

KW noodle; animal feed; starch separation; maize processing; malting;

KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX *Triticum aestivum* cultivar Estica.  
 OS Synthetic.  
 XX WO200198474-A1.  
 PM 27-DEC-2001.  
 PD 21-JUN-2001; 2001WO-BE00106.  
 PF 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX (LEUV-) LEEVEN RES & DEV.  
 FA Delcour J, Debysse W, Gebruers K, Goesaert H, Flerens K, Robben J;  
 FI Van Campenhout S;  
 FI WPI: 2002-114579/15.  
 DR N-PSDB: ABK13690.  
 XX Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprising using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX Example 11: Fig 21: 127pp: English.  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syripping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents the sequence surrounding the N-terminus of wheat  
 CC TAXI-I (T. aestivum L endoxylanase inhibitor) which was expressed in  
 CC E. coli.  
 CC  
 XX  
 SQ Sequence 22 AA:  
 Query Match 78.9%; Score 90; DB 23; Length 22;  
 Best Local Similarity 94.7%; Pred No. 2.6e-06;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 LPVPAPVTKDPATSLYTIP 19

Db III |||||  
 4 LPVPAPVTKDPATSLYTIP 22  
 RESULT 9  
 AAU75807  
 ID AAU75807 standard; peptide: 22 AA.  
 AC AAU75807;  
 XX 23-APR-2002 (first entry)  
 DT  
 DE Wheat L endoxylanase inhibitor, TAXI II, N-terminus #1.  
 XX  
 DE  
 KW Wheat; TAXI-II: L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syripping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS *Triticum aestivum*.  
 XX  
 PN WO200198474-A1.  
 XX 27-DEC-2001.  
 PD 21-JUN-2001; 2001WO-BE00106.  
 PF 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX (LEUV-) LEEVEN RES & DEV.  
 PA Delcour J, Debysse W, Gebruers K, Goesaert H, Flerens K, Robben J;  
 PI Van Campenhout S;  
 PI WPI: 2002-114579/15.  
 DR Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 XX beta-glucanolytic enzymes comprising using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX Claim 127; Page 9; 127pp: English.  
 PS  
 XX The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syripping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as

CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
CC biscuits, pasta and noodles, animal feed stuff, improving the production  
CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
CC starch separation and production, improving maize processing,  
CC plant disease resistance and nutraceutical and/or pharmaceutical  
CC applications, improving paper and pulp technologies. The present  
CC sequence is the N-terminus of wheat TxAI-II (T. aestivum L. endoxylanase  
CC inhibitor) from a 40-43kDa isolate of the protein.  
XX  
SQ Sequence 22 AA:  
Query Match 77.2%; Score 88; DB 23; Length 22;  
Best Local Similarity 90.0%; Pred. No. 5.1e-06;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 LPVPAPVTKDPATSLTYTTF 20  
Db 3 LPVLAPVTKDPATSLTYTTF 22  
RESULT 10  
AA93764  
ID AA93764 standard; peptide: 38 AA.  
XX  
AC AA93764;  
XX  
DT 03-OCT-2000 (first entry)  
XX  
DE Amino acid sequence derived from an endo-beta-1,4-xylanase inhibitor.  
XX  
KW Endo-beta-1,4-xylanase inhibitor; wheat flour; xylanase; bakery product;  
KW dough; dough preparation.  
XX  
OS Triticum sp.  
XX  
FH Key Location/Qualifiers  
FH Misc-difference 26 /note= "unspecified amino acid"  
FT Misc-difference 31 /note= "unspecified amino acid"  
FT Misc-difference 32 /note= "unspecified amino acid"  
FT Misc-difference 38 /note= "unspecified amino acid"  
XX  
PN WO200039289-A2.  
XX  
PD 06-JUL-2000.  
XX  
PF 17-DEC-1999; 99WO-1B02071.  
XX  
PR 23-DEC-1998; 98GB-0028599.  
PR 06-APR-1999; 98GB-0007805.  
PR 15-APR-1999; 99GB-0008645.  
XX  
PA (DANI-) DANISCO AS.  
XX  
PI Sidsesen O, Sorensen JF;  
XX  
DR WPI: 2000-465744/40.  
XX  
PT Mutant xylanase protein identified using xylanase inhibitor useful for  
PT preparing non-sticky dough for bakery products  
XX  
PS Disclosure: Page 106; 112pp; English.  
XX  
CC The present sequence is derived from an endo-beta-1,4-xylanase  
CC inhibitor. The protein is obtained from wheat flour. The specification  
CC also describes a mutant xylanase protein. The xylanase is useful for  
CC preparing a foodstuff, preferably a bakery product or a substance  
CC (e.g. a dough) for making the bakery product. Wild type xylanase or  
CC mutant xylanase is useful for preparing a dough that is less sticky  
CC than a dough comprising a fungal xylanase. The xylanase inhibitor is

CC useful for screening high degree resistance xylanases for dough  
CC preparation. The xylanase is also useful for preparing a non-sticky  
CC dough. A combination of xylanase and the inhibitor is useful for  
CC calibrating and/or determining the quantity of inhibitor in a wheat  
CC flour sample.  
XX  
SQ Sequence 38 AA:  
Query Match 71.9%; Score 82; DB 21; Length 38;  
Best Local Similarity 80.0%; Pred. No. 6.9e-05;  
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 2 PVPAPVTKDPATSLTYTTF 21  
Db 2 PPLAPVTEAPATSLTYTTF 21  
RESULT 11  
AAU75808  
ID AAU75808 standard; peptide: 22 AA.  
XX  
AC AAU75808;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE Barley L endoxylanase inhibitor, HvXI L, N-terminus #1.  
XX  
KM Barley; HvXI-L; L endoxylanase inhibitor; cellulosytic enzyme inhibitor;  
KM xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
KM immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
KM straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
KW noodle; animal feed; starch separation; maize processing; malting;  
KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
XX  
OS Hordeum vulgare.  
XX  
FH Key Location/Qualifiers  
FH Misc-difference 21 /label= OTHER  
FT Misc-difference 22 /note= "Other- Any amino acid, preferably Pro"  
FT Misc-difference 22 /label= OTHER  
FT /note= "Other- Any amino acid, preferably Phe"  
XX  
PN WO200198474-A1.  
XX  
PD 27-DEC-2001.  
XX  
PF 21-JUN-2001; 2001WO-BE00106.  
XX  
PR 22-JUN-2000; 2000GB-0015296.  
PR 25-JAN-2001; 2001GB-0002018.  
PR 26-JAN-2001; 2001GB-0002194.  
PR 16-MAR-2001; 2001GB-0006564.  
PR 21-MAY-2001; 2001GB-0012328.  
XX  
PA (LEUV-) LEUVEN RES & DEV.  
XX  
PI Delcourt J, Debysse W, Gebuere K, Goesaert H, Pierens K, Robben J;  
PI Van Campenhout S;  
XX  
DR WPI: 2002-114579/15.  
XX  
PT Separating and/or isolating inhibitors of cellulosytic, xylanolytic, or  
PT beta-glucanolytic enzymes comprises using endoxylanases during  
PT screening for inhibition activity or affinity chromatography with  
PT immobilised enzymes  
XX  
PS Claim 127; Page 9; 127pp; English.  
XX  
CC The invention relates to separating and/or isolating inhibitors of  
CC cellulosytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
CC screening the inhibition activity by using two or more enzymes during the

CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruiping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is the N-terminus of barley HvXI-I (L endoxylanase  
 CC inhibitor) from a 40-43kDa isolate of the protein.

XX Sequence 22 AA:

Query Match 65.8%; Score 75; DB 23; Length 22;  
 Best Local Similarity 88.9%; Pred. No. 0.00039;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 LPVPAPVTKDPATSLTYT 18  
 ||| ||||| ||||| |||||  
 Db 3 LPVLAPVTKDAATSLTYT 20

RESULT: 12

AAU75823 standard; Protein: 74 AA.

AC AAU75823;

DT 23-APR-2002 (first entry)

DE Maize L endoxylanase inhibitor, partial sequence ZMXI-01.

KW Maize; ZMXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruiping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

XX Zea mays.

PN WO200198474-A1.

PD 27-DEC-2001.

PF 21-JUN-2001; 2001WO-BE00106.

PR 22-JUN-2000; 2000GB-0015296.

PR 25-JAN-2001; 2001GB-0002018.

PR 16-MAR-2001; 2001GB-0002194.

PR 21-MAY-2001; 2001GB-0012328.

PA (LEUV-) LEUVEN RES & DEV.

XX Delcours J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR WPI, 2002-114579/15.  
 DR N-PSDB; ABK13682.  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 PS Claim 127; Page 62; 127pp; English.

CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruiping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial rice ZMXI-I (Z. mays L endoxylanase  
 CC inhibitor).

SO Sequence 74 AA:

Query Match 47.4%; Score 54; DB 23; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ATSLTYIPPH 21  
 ||||| ||||| ||||| |||||  
 Db 1 ATSLTYIPPH 10

RESULT: 13

AAU75819 standard; Protein: 287 AA.

AC AAU75819;

DT 23-APR-2002 (first entry)

DE Wheat L endoxylanase inhibitor, partial sequence TAXI-III.

KW Wheat; TAXI-III; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruiping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;

KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX Triticum aestivum cultivar Soissons.  
 XX WO200198474-A1.  
 XX 27-DEC-2001.  
 XX 21-JUN-2001; 2001WO-BE00106.  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 23-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX (LEUV-) LEUVEN RES & DEV.  
 PA Delcour J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 PI WPI: 2002-114579/15.  
 DR N-PSDB: ABK13676.  
 DR Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 XX beta-glucanolytic enzymes comprising using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX Claim 127; Page 58; 127pp; English.  
 XX The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arbinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malling of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial wheat TAXI-III (T. aestivum L endoxylanase  
 CC inhibitor).  
 XX  
 XX Sequence 287 AA:  
 SO  
 Query Match 47.4%; Score 54; DB 23; Length 287;  
 Best Local Similarity 100.0%; Pred. No. 7.1;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 12 ATSLYTIPIPH 21  
 |||||||||  
 Db 1 ATSLYTIPIPH 10

RESULT 14  
 AAU75815  
 ID AAU75815 standard; Protein: 370 AA.  
 XX  
 XX AAU75815;  
 AC  
 AC 23-APR-2002 (first entry)  
 DT  
 DT  
 DE Wheat L endoxylanase inhibitor, TAXI I, partial sequence TAXI-I.0.  
 XX  
 XX Wheat; TAXI-I: L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malling;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 XX Triticum aestivum cultivar Soissons.  
 OS  
 OS WO200198474-A1.  
 PN  
 PN 27-DEC-2001.  
 PD  
 PD 21-JUN-2001; 2001WO-BE00106.  
 XX  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 23-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX (LEUV-) LEUVEN RES & DEV.  
 PA Delcour J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 PI WPI: 2002-114579/15.  
 DR N-PSDB: ABK13672.  
 DR Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 XX beta-glucanolytic enzymes comprising using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX Claim 127; Page 57; 127pp; English.  
 PS The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arbinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malling of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:11:22 ; Search time 8.53846 Seconds  
(without alignments)  
236.439 Million cell updates/sec

Title: US-09-869-155-15  
Perfect score: 114  
Sequence: 1 LPVPAVPYTKDPATSLYTRPFH 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51.5	45.2	100	2	T18020
2	50	43.9	3498	2	T22330
3	49	43.0	532	2	S74453
4	49	43.0	1047	2	A55617
5	48.5	42.5	402	2	F96640
6	48	42.1	273	2	G82049
7	48	42.1	434	2	G86163
8	47	41.2	336	1	S75947
9	47	41.2	336	2	T15837
10	47	41.2	750	2	AB0708
11	47	41.2	1075	2	G59434
12	47	41.2	1199	2	T13946
13	46	40.4	151	2	S10084
14	46	40.4	186	2	A70574
15	46	40.4	760	2	T06291
16	46	40.4	1299	2	T47182
17	46	40.4	1473	2	A35186
18	46	40.4	13055	2	T16580
19	45	39.5	241	2	T48006
20	45	39.5	275	2	T51437
21	45	39.5	432	2	AE1285
22	45	39.5	435	2	E72563
23	45	39.5	440	2	JC2065
24	45	39.5	461	2	S45568
25	45	39.5	470	2	JC2064
26	45	39.5	498	2	S45567
27	45	39.5	509	2	S45565
28	45	39.5	522	2	S09996
29	45	39.5	597	2	AH2351

30	45	39.5	753	2	A39129	catalase (EC 1.11.
31	45	39.5	753	2	B85782	catalase, hydrop
32	45	39.5	753	2	F90933	catalase, HP11 (imp
33	45	39.5	1048	2	T31425	C-terminal domain-
34	44.5	39.0	532	2	JT0530	muscarinic acetylch
35	44.5	39.0	1196	2	T24222	hypothetical prote
36	44	38.6	187	2	S38036	hypothetical prote
37	44	38.6	357	2	PC4293	nuclear factor 1 f
38	44	38.6	390	1	Q08E77	glycoprotein I pre
39	44	38.6	416	1	A31959	lysosome-associate
40	44	38.6	426	2	T20265	hypothetical prote
41	44	38.6	460	2	T45968	hypothetical prote
42	44	38.6	504	2	A57215	glial cells missin
43	44	38.6	509	2	JC5428	nuclear factor 1 f
44	44	38.6	532	2	B36596	nuclear factor I -
45	44	38.6	549	1	N0EC	glucose-6-phosphat

## ALIGNMENTS

```

RESULT 1
T18020
hypothetical protein a518r - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18020
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T18020
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-100 <GRA>
A:Cross-References: EMBL:U42580; NID:g4028896; PIDN:AAC36885.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Note: a518r

Query Match      45.2%; Score 51.5; DB 2; Length 100;
Best Local Similarity 45.8%; Pred. No. 1.2;
Matches 11; Conservative 3; Mismatches 7; Indels 3; Gaps 1;

OY      1 LPVPA---PYTKDPATSLYTRPFH 21
DB      17 VPPPAVEKPYQGHPAVWMTLPFH 40

RESULT 2
T22330
hypothetical protein F47A4.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T22330
R:Mortimore, B.
submitted to the EMBL Data Library, June 1995
A:Reference number: Z19549
A:Accession: T22330
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3498 <WIL>
A:Cross-References: EMBL:Z49888; PIDN:CAA90064.1; GSPDB:GN00028; CESP:F47A4.2
A:Experimental source: clone F47A4
C:Genetics:
A:Gene: CESP:F47A4.2
A:Map position: X
A:Introns: 36/3; 72/3; 147/2; 228/3; 261/2; 320/2; 847/3; 881/1; 1317/2; 1500/3; 1653

Query Match      43.9%; Score 50; DB 2; Length 3498;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY      2 PVPAPVTKDPAT 13

```

Db 2305 PKPAPYTKRSPAT 2316

# RESULT 3

hypothetical protein slr1484 - *Synechocystis* sp. (strain PCC 6803)  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C:Accession: S74453  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S74453

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-532 <KAN>

A:Cross-references: EMBL:D90899; GB:AB001339; NID:91651650; PIDN:BAAL6605.1; PID:dl01733  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

## Query Match

Best Local Similarity 43.0%; Score 49; DB 2; Length 532;  
Matches 10; Conservative 5; Mismatches 4; Indels 6; Gaps 1;

OY 2 PVPAPYTKDPATSLTYTIP 20

Db 161 PLNPVTESPAITOPAGSAFLRP 185

# RESULT 4

masquerade precursor - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 09-Mar-1996 #sequence\_revision 09-Mar-1996 #text\_change 21-Jul-2000

C:Accession: A55617

R:Murugasu-Oel, B.; Rodrigues, V.; Yang, X.; Chia, W.

Genes Dev. 9, 139-154, 1995

A:Title: Masquerade: a novel secreted serine protease-like molecule is required for some

A:Reference number: A55617; MUID:95154720; PMID:7851790

A:Accession: A55617

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1047 <RES>

A:Cross-references: EMBL:U18130; NID:g665544; PIDN:ANCA6512.1; PID:g665545

A:Comment: This secreted protein, unlikely to function as a serine proteinase, is proces

A:Genes: mas

A:Cross-references: FlyBase:FBgn0011653

C:Superfamily: trypsin homology

F:803-1038/Domain: trypsin homology <TRY>

Query Match 43.0%; Score 49; DB 2; Length 1047;  
Best Local Similarity 61.1%; Pred. No. 43;  
Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 PVPAPYTKDPATSLTYTIP 19

Db 141 PKKPOTKRPAATSTTKP 158

# RESULT 5

hypothetical protein T25B24.11 [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C:Accession: F96640

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Matzla

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: F96640

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-402 <STO>

A:Cross-references: GB:AE005173; NID:94585881; PIDN:AA25554.1; GSPDB:GN00141

C:Genetics:

A:Gene: T25B24.11

C:Superfamily: caffeoyl-CoA 3-O-methyltransferase

## Query Match

Best Local Similarity 42.5%; Score 48.5; DB 2; Length 402;  
Matches 8; Conservative 8; Mismatches 3; Indels 1; Gaps 1;

OY 1 LPVPAPYTKDPATSLTYTIP 20

Db 14 MPVPVTEEP-ISTMSLPY 32

# RESULT 6

serine acetyltransferase VC2649 [imported] - *Vibrio cholerae* (strain N16961 serogroup

C:Species: *Vibrio cholerae*

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: G82049

R:Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers

L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: G82049

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-273 <HEI>

A:Cross-references: GB:AE004331; GB:AE003852; NID:g9657236; PIDN:AAFP5790.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2649

A:Map position: 1

C:Superfamily: serine acetyltransferase; serine acetyltransferase homology

## Query Match

Best Local Similarity 42.1%; Score 48; DB 2; Length 273;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 5 APYTKDPATSLTYTIP 19

Db 87 ATVNRDPAVMSWSP 101

# RESULT 7

hypothetical protein F15K9.16 - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C:Accession: G86163

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marcia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo



ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G86163

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-434 <STO>

A:Cross-References: GB:AE005172; NID:g3850580; PIDN:AACT2120.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: conglutin gamma

Query Match

Best Local Similarity 42.1%; Score 48; DB 2; Length 434;  
 Pred. No. 22;

Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 6 PVTKDPATSLYTI 17

DB 35 PVTKDPSTLOYT 46

RESULT 8

hypothetical protein - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.

A:Variate: PCC 6803

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C:Accession: S75947

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

S.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S75947

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-336 <KAN>

A:Cross-References: EMBL:D64006; GB:AB001339; NID:g1001291; PID:g1001307

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: conserved hypothetical protein ylnf

Query Match

Best Local Similarity 41.2%; Score 47; DB 1; Length 336;  
 Pred. No. 23;

Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 3 VPAVPTKDPATSLYTIPT 21

DB 6 VPAVPSLFPETLEPLPYH 24

RESULT 9

hypothetical protein C54D2.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T15837

R:Minx, P.

submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid C54D2.

A:Reference number: Z18415

A:Accession: T15837

A:Status: preliminary; translated from GB/EMBL/DDJ

A:Molecule type: DNA

A:Residues: 1-336 <MIN>

A:Cross-References: EMBL:U37548; NID:g1017804; PID:g1017808; PIDN:AAAY9200.1; CESP:C54D2

C:Genetics:

A:Gene: CESP:C54D2.1

A:Introns: 31/1; 87/1; 116/1; 150/2; 221/3; 295/3

Query Match

Best Local Similarity 41.2%; Score 47; DB 2; Length 336;  
 Pred. No. 23;

Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1 LPVAPVTKDPATSLYTI 19

DB 116 VPVVPEVPTATPATVAPTP 134

RESULT 10

AB0708

catalase (EC 1.11.1.6) - *Salmonella enterica* subsp. *enterica* serovar Typh1 (strain CT

C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh1

A:Note: This species has also been called *Salmonella typhi*

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 17-May-2002

C:Accession: AB0708

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se

A:Reference number: AB0502; PMID:11677608

A:Accession: AB0708

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-750 <PAR>

A:Cross-References: GB:AL513382; PIDN:CAD02035.1; PID:g16502872; GSPDB:GN00176

C:Genetics:

A:Gene: SRY1793

C:Superfamily: catalase

C:Keywords: oxidoreductase

Query Match

Best Local Similarity 41.2%; Score 47; DB 2; Length 750;  
 Pred. No. 58;

Matches 10; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

OY 4 PAP---VTKDPATSLYTI 19

DB 572 PPPVNGKPKDPALSLYAVP 591

RESULT 11

G59434

KIAA0411 protein [imported] - human

C:Species: Homo sapiens (man)

C>Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 19-Jul-2002

C:Accession: G59434

R:Turner, A.C.; Zambrowicz, B.; Nehls, M.; Friedrich, G.A.; Sands, A.T.

patent application, Lexicon Genetics Incorporated (US), WO 0075320-A, December 2000

A:Description: Human genes and proteins encoded thereby.

A:Reference number: G59434

A:Accession: G59434

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1075 <TUR>

A:Cross-References: GB:CAZ22407; PID:g12309846; PIDN:CAZ22407.1

Query Match

Best Local Similarity 41.2%; Score 47; DB 2; Length 1075;  
 Pred. No. 89;

Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 4 PAPVTKDPATSLYTI 18

DB 978 PGVPSSEPASPLHTI 992

RESULT 12

T13946

probable adaptor-related protein complex AP-3, delta chain - mouse

N:Alternate names: bovine leukemia virus receptor homolog

C:Species: Mus musculus (house mouse)

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000

C:Accession: T13946

R:Suzuki, T.; Ikeda, H.

J. Virol. 72, 593-599, 1998

Search completed: May 20, 2003, 16:37:00  
Job time : 10.5385 secs

Query Match	40.4%	Score 46;	DB 2;	Length 760;
Best Local Similarity	44.4%	Pred. No. 84;		
Matches	8;	Conservative	3;	Mismatches 7;
				Indels 0;
				Gaps 0;
QY	2	PVPAPVTKDPATSLYTIP	19	
Db	593	PPPTPVSSPPPTPVVSSP	610	

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK protein - protein search, using sw model

Run on: May 20, 2003, 16:05:42 ; Search time 5.65385 Seconds  
(without alignments)  
154.055 Million cell updates/sec

Title: US-09-869-155-15  
Perfect score: 114  
Sequence: 1.LPVPAVTKDPATSLYTRPFH 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	43.9	349	1	IRF2_MOUSE
2	48	42.1	179	1	RBS1_FRIG
3	48	42.1	179	1	RBS3_FRIG
4	47	41.2	650	1	Y411_HUMAN
5	46	40.4	151	1	WMT5_WHEAT
6	46	40.4	555	1	DAB1_HUMAN
7	46	40.4	555	1	DAB1_MACR
8	46	40.4	806	1	MXO7_MOUSE
9	45	39.5	328	1	HXCD_MOUSE
10	45	39.5	330	1	HXCD_MOUSE
11	45	39.5	470	1	P26_RANCA
12	45	39.5	509	1	NR1A_HUMAN
13	45	39.5	522	1	NR1A_CHICK
14	45	39.5	753	1	CATE_ECOLI
15	45	39.5	1048	1	SRA4_RAT
16	44.5	39.0	532	1	ACM5_HUMAN
17	44	38.6	187	1	YKT9_YEAST
18	44	38.6	390	1	VGLI_HSV1
19	44	38.6	416	1	LMP1_HUMAN
20	44	38.6	509	1	NR1A_RAT
21	44	38.6	532	1	NR1A_MOUSE
22	44	38.6	549	1	G6P1_ECOLI
23	44	38.6	562	1	LCB2_KLUGA
24	44	38.6	669	1	C101_RAT
25	44	38.6	1844	1	POLR_TYMW
26	44	38.6	1844	1	POLR_TYMW
27	44	38.6	1844	1	POLR_TYMW
28	43.5	38.2	878	1	IL3B_MOUSE
29	43	37.7	116	1	U6B1_HICMA
30	43	37.7	116	1	U6B1_HICMA
31	43	37.7	116	1	U6B1_HICMA
32	43	37.7	116	1	U6B1_HICMA
33	43	37.7	116	1	U6B1_HICMA

34	43	37.7	1157	1	SRA4_HUMAN	O95104 homo sapien
35	43	37.7	1253	1	SHK2_HUMAN	O949X8 homo sapien
36	43	37.7	1324	1	POL2_GCMV	P13026 hungarian g
37	43	37.7	1474	1	SHK2_RAT	O94974 rattus norv
38	43	37.7	3591	1	FNAB_BOPE	P12255 bordelella
39	42.5	37.3	215	1	HIS5_DEIRA	O91X89 deinococcus
40	42.5	37.3	432	1	ST11_XENLA	O91604 xenopus lae
41	42.5	37.3	601	1	DNM1_MYGE	P47442 mycoplasma
42	42.5	37.3	906	1	CR1_HUMAN	P22681 homo sapien
43	42	36.8	182	1	YR31_CAEBL	O10958 caenorhabdi
44	42	36.8	273	1	ROCI_NICSY	O08935 nicotiana s
45	42	36.8	401	1	NNOS_DROME	P25724 drosophila

## ALIGNMENTS

RESULT 1  
ID IRF2\_MOUSE STANDARD: PRT: 349 AA.  
AC P23906:  
DF 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Interferon regulatory factor 2 (IRF-2).  
GN IRF2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89354547; PubMed=2475256;  
RA Harada H., Fujita T., Miyamoto M., Kimura Y., Maruyama M.,  
RA Furia A., Miyata T., Taniguchi T.;  
RT "Structurally similar but functionally distinct factors, IRF-1 and  
RT IRF-2, bind to the same regulatory elements of IFN and IFN-inducible  
RT genes";  
RT Cell 58:729-739(1989).  
RN [2]  
RP STRUCTURE BY NMR OF 2-113.  
RX MEDLINE=98230747; PubMed=9562558;  
RA Furi J., Uegaki K., Yamazaki T., Shirakawa M., Swindells M.B.,  
RA Harada H., Taniguchi T., Kyogoku Y.;  
RT "Solution structure of the IRF-2 DNA-binding domain: a novel subgroup  
RT of the winged helix-turn-helix family.";  
RT Structure 6:491-500(1998).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1-113.  
RX MEDLINE=99417584; PubMed=10487755;  
RA Fujii Y., Shimizu T., Kusumoto M., Kyogoku Y., Taniguchi T.,  
RA Hakoshima T.;  
RT "Crystal structure of an IRF-DNA complex reveals novel DNA recognition  
RT and cooperative binding to a tandem repeat of core sequences.";  
RT EMBO J. 18:5028-5041(1999).  
CC -!- FUNCTION: SPECIFICALLY BINDS TO THE UPSTREAM REGULATORY REGION OF  
CC TYPE I IFN AND IFN-INDUCIBLE MHC CLASS I GENES (THE INTERFERON  
CC CONSENSUS SEQUENCE (ICS1)) AND REPRESSES THOSE GENES.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- INDUCTION: BY VIRUSES AND IFN.  
CC -!- SIMILARITY: BELONGS TO THE IRF FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
CC -----  
CC EMBL: J03168; AAA93333.1; -  
CC PIR: A32828; A32828.  
CC DR PDB: 1IRF; 28-JAN-96.

DR DB: 21RF; 08-OCT-99.  
 DR DB: 11RG; 18-MAR-98.  
 DR TRANSFAC; T00425; -.  
 DR MGD; MGI:96591; 1rf2.  
 DR InterPro; IPR001346; 1RF.  
 DR Pfam; PF00605; 1RF; 1.  
 DR PRINTS; PR00267; INTERNRGCT.  
 DR ProDom; PD002355; 1RF; 1.  
 DR SMART; SM00348; 1RF; 1.  
 DR PROSITE; PS00601; 1RF; 1.  
 KW Transcription regulation; DNA-binding; Repressor; Nuclear protein;  
 KW Interference induction; 3D-structure.  
 FT DNA\_BIND 7 109 TRYPTOPHAN PENTAD REPEAT.  
 SQ SEQUENCE 349 AA; 39452 MW; 8738B082FB40FB11 CRC64;

Query Match 43.9%; Score 50; DB 1; Length 349;  
 Best Local Similarity 64.3%; Pred. No. 5; Mismatches 2; Indels 0; Gaps 0;  
 Matches 9; Conservative 3;

OY 1 LPVPAPVTKDPATS 14  
 ||:||||| 1::1  
 Db 311 LPVPAPVTPPSSS 324

RESULT 2  
 RBS1\_FRIAG STANDARD; PRT; 179 AA.  
 AC 024634:

DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ribulose biphosphate carboxylase small 1/4, chloroplast  
 DE precursor (EC 4.1.1.39) (Rubisco small subunit 1/4).

GN RBCS4.  
 OS Fritillaria agrestis.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Liliaceae;  
 OC Fritillaria.  
 OC NCBI\_TaxID=64177;

SEQUENCE FROM N.A.  
 RA Panico E., Baysdorfer C.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
 D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 ACTIVE SITE.

CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) -> 2 3-  
 phospho-D-glycerate.  
 CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) ->  
 3-phospho-D-glycerate + 2-phosphoglycolate.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.

CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF031543; AAB86853.1; -.  
 DR EMBL; AF024572; AAB84179.1; -.  
 DR InterPro; IPR000894; Rubisco\_small.  
 DR Pfam; PF00101; Rubisco\_small; 1.  
 DR PRINTS; PR00152; RUBISCO\_SMALL.  
 DR ProDom; PD000290; Rubisco\_small; 1.  
 KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
 KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;

KW Multigene family.  
 FT TRANSIT 1 59  
 FT CHAIN 60 179  
 FT SEQUENCE 179 AA; 19620 MW; 1695775BD611A765 CRC64;

Query Match 42.1%; Score 48; DB 1; Length 179;  
 Best Local Similarity 64.3%; Pred. No. 4.8; Mismatches 3; Indels 0; Gaps 0;  
 Matches 9; Conservative 2;

OY 6 PVTKDPATSLYTP 19  
 |||:|||| 1:1  
 Db 38 PVTKDPATGLSTLP 51

RESULT 3  
 RBS3\_FRIAG STANDARD; PRT; 179 AA.  
 ID RBS3\_FRIAG  
 AC 022573:

DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ribulose biphosphate carboxylase small chain 3, chloroplast precursor  
 DE (EC 4.1.1.39) (Rubisco small subunit 3).

GN RBCS3.  
 OS Fritillaria agrestis.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Liliaceae;  
 OC Fritillaria.  
 OC NCBI\_TaxID=64177;

SEQUENCE FROM N.A.  
 RA Panico E., Baysdorfer C.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
 D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 ACTIVE SITE.

CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) -> 2 3-  
 phospho-D-glycerate.  
 CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) ->  
 3-phospho-D-glycerate + 2-phosphoglycolate.

CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.

CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF024574; AAB84181.1; -.  
 DR InterPro; IPR000894; Rubisco\_small.  
 DR Pfam; PF00101; Rubisco\_small; 1.  
 DR PRINTS; PR00152; RUBISCO\_SMALL.  
 DR ProDom; PD000290; Rubisco\_small; 1.  
 KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
 KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;  
 KW Multigene family.  
 FT TRANSIT 1 59  
 FT CHAIN 60 179  
 FT SEQUENCE 179 AA; 19615 MW; D2C68A6CC0E0BF66 CRC64;

Query Match 42.1%; Score 48; DB 1; Length 179;  
 Best Local Similarity 64.3%; Pred. No. 4.8; Mismatches 3; Indels 0; Gaps 0;  
 Matches 9; Conservative 2;

Oy 6 PVTKDPATSLYTP 19  
 Db 38 PVTOKPATGLSTLP 51

## RESULT 4

Y411\_HUMAN STANDARD; PRT; 650 AA.  
 AC 043295;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUN-1998 (Rel. 36, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 GN KIAA0411.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=9811655; PubMed=9455477;  
 RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VIII.  
 RT 78 new cDNA clones from brain which code for large proteins in  
 RT vitro."  
 RL DNA Res. 4:307-313(1997).  
 CC -1- SIMILARITY: CONTAINS 1 RHO-GAP DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB007871; BAA24841.1; -  
 DR HSSP: O60631; IGRO.  
 DR InterPro: IPR00198; RHO GAP.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00018; SH3; 1.  
 DR Pfam: PF00620; RHO GAP; 1.  
 DR ProDom: PD00066; SH3; 1.  
 DR SMART: SM00324; RHO GAP; 1.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 KW Hypothetical protein; GTPase activation; SH3 domain.  
 FT NON\_TER 1 1  
 FT DOMAIN 71 214 RHO-GAP.  
 FT DOMAIN 296 365 SH3.  
 SO SEQUENCE 650 AA; 71863 MW; 7175AA9CCF259907 CRC64;  
 Query Match 41.2%; Score 47; DB 1; Length 650;  
 Best Local Similarity 53.3%; Pred. No. 26;  
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 Oy 4 PAPVTKDPATSLYTT 18  
 Db 553 PGVSSSEBASPLHTI 567

## RESULT 5

Y411\_HUMAN STANDARD; PRT; 151 AA.  
 AC 043295;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 16 kDa protein in middle repetitive insertion sequence  
 DE WSI.

OS Triticum aestivum (Wheat).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 CC Triticeae; Triticum.  
 CC NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Chinese Spring;  
 RX MEDLINE=89364725; PubMed=2549380;  
 RA Martienssen R.A., Paulcombe D.C.;  
 RT "An unusual wheat insertion sequence (WS1) lies upstream of an  
 RT alpha-amylase gene in hexaploid wheat, and carries a 'minisatellite'  
 RT array."  
 RL Mol. Genet. 217:401-410(1989).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X15870; CAA33880.1; -  
 DR PIR: S10084; S10084.  
 KW Transposable element; Hypothetical protein.  
 SO SEQUENCE 151 AA; 15950 MW; 9BAB30BD31EC6742 CRC64;

Query Match 40.4%; Score 46; DB 1; Length 151;  
 Best Local Similarity 47.4%; Pred. No. 7.7;  
 Matches 9; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Oy 2 PVPAPVTKDPATSLYTP 20  
 Db 64 PPPPTSPSISLPPFR 82

## RESULT 6

DAB1\_HUMAN STANDARD; PRT; 555 AA.  
 AC 073553; Q9NYA8;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Disabled homolog 1.  
 GN DAB1.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99009343; PubMed=9790777;  
 RA Lambert de Rouvroit C., Goffinet A.M.;  
 RT "Cloning of human DAB1 and mapping to chromosome 1p31-p32."  
 RL Genomics 53:246-247(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Fazili Z., Sun W., Xu X.-X.;  
 RT "Aberant disabled-1 expression in tumors."  
 RT Submitted (May-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Adaptor molecule functioning in neural development (By  
 CC similarity).  
 CC -1- SUBUNIT: Associates with the SH2 domains of Src, Fyn and Abl (By  
 CC similarity).  
 CC -1- DOMAIN: THE PID DOMAIN SPECIFICALLY BINDS TO THE ASN-PRO-XAA-  
 CC TYR(P) MOTIF FOUND IN MANY TYROSINE-PHOSPHORYLATED PROTEINS.  
 CC -1- PTM: Phosphorylated on tyrosine (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 PID DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its



```

KW Phosphorylation.
FT DOMAIN 55 347 PROTEIN KINASE.
FT DOMAIN 434 465 PRO-RICH 1.
FT DOMAIN 521 524 POLY-ARG.
FT DOMAIN 578 700 PRO-RICH 2.
FT NP_BIND 61 69 ATP (BY SIMILARITY).
FT BINDING 84 84 ATP (BY SIMILARITY).
FT ACT_SITE 182 182 BY SIMILARITY.
FT MOD_RES 219 219 PHOSPHORYLATION (ACTIVATES THE KINASE)
FT MOD_RES 221 221 (BY SIMILARITY).
FT MOD_RES 221 221 PHOSPHORYLATION (ACTIVATES THE KINASE)
FT MOD_RES 221 221 (BY SIMILARITY).
SO SEQUENCE 806 AA: 87732 MW: E7CC41C4BDE0633 CRC64;

Query Match
Best Local Similarity 57.1%; Score 46; DB 1; Length 806;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 PVPAPVTKDPATSL 15
DB 640 PIPAPLQTAPTSLS 653

RESULT 9
HXCD_MOUSE STANDARD: PRT; 328 AA.
ID HXCD_MOUSE
AC P50207; 092017;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Homeobox protein Hox-C13.
GN HOXC13 OR HOXC-13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN NCBI_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=FVB;
RA Tkatchenko A.V., Visconti R.P., Shang L., Papenbrock T., Ito T.,
RA Ogawa M., Awgulewitsch A.;
RT "Hyperproliferation of keratinocytes, alopecia and Ichthyosis-like
RT syndrome in transgenic mice overexpressing Hoxc13."
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 272-309 FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=95048380; PubMed=7959778;
RA Bradshaw M.S., Ruddle F.H.;
RT "Identification of the murine Hox-c12 and Hox-c13 homeoboxes on yeast
RT artificial chromosomes."
RL Genomics 22:234-236(1994).
CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ABD-B HOMEOBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL; AF193796; AAL09298.1; -
CC EMBL; U04838; AAL20229.1; -
CC TRANSPAC; T03354; -
CC MGI; 99560; Hoxc13.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; homeobox.2.
CC ProDom; PD000010; Homeobox.1.

```

```

DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEOBOX_1; 1.
DR PROSITE; PS00071; HOMEOBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DNA_BIND 258 317 HOMEOBOX.
FT DOMAIN 27 60 GLY-RICH.
SO SEQUENCE 328 AA: 35192 MW: AE363563B5AEF776 CRC64;

Query Match
Best Local Similarity 39.5%; Score 45; DB 1; Length 328;
Matches 10; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 2 PVPAPVTKDPATSLTYTIPF 20
DB 105 PPPAPVTSSATILGYGYPF 123

RESULT 10
HXCD_HUMAN STANDARD: PRT; 330 AA.
ID HXCD_HUMAN
AC P31276; Q9NRYD5; Q9NR24; Q9GJ32;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Homeobox protein Hox-C13 (Hox-3G).
GN HOXC13 OR HOX3G.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RX MEDLINE=20296799; PubMed=10835276;
RA de Stanchina E., Gabellini D., Norio P., Giacca M., Peverali F.A.,
RA Riva S., Falaschi A., Diamonti G.;
RT "Selection of homeotic proteins for binding to a human DNA replication
RT origin."
RL J. Mol. Biol. 299:667-680(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Li S., Zhang B., Li X.;
RT "A homeobox protein interacting with promoter region of p21WAF1/Cip1
RT gene is identical to HOX-C13 (Hox 3G).";
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strusberg R.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 260-325 FROM N.A.
RX MEDLINE=90098876; PubMed=2574852;
RA Acampora D., D'Esposito M., Faiella A., Pannese M., Migliaccio E.,
RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
RT "The human HOX gene family."
RL Nucleic Acids Res. 17:10385-10402(1989).
RN [5]
RP SEQUENCE OF 185-330 FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagstsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Yamane S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDD human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ABD-B HOMEOBOX FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF255676; AAF67760.1; -
DR EMBL: AF263466; AAF73439.1; -
DR EMBL: BC002754; AAH02754.1; -
DR EMBL: AK024027; BAB14786.1; -
DR PIR: S14934; S14934.
DR HSSP: P14653; 1B72.
DR TRANSPAC: T03330; -
DR Genew: HGNC:5125; HOKC13.
DR MIM: 142976; -
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox.1.
DR ProDom: PD000010; Homeobox.1.
DR SMART: SM00389; HOX.1.
DR PROSITE: PSS00027; HOMEBOX_1; 1.
DR PROSITE: PSS0071; HOMEBOX_2; 1.
DR Homeobox: DNA-binding; Developmental protein; Nuclear protein;
DR Transcription regulation; Polymorphism.
FT DNA_BIND 260 319 HOMEBOX.
FT DOMAIN 27 61 GLY-RICH.
FT VARIANT 50 50 S->I (IN DBSNP:1867298).
FT CONFLICT 95 95 D->E (IN REF.1).
FT SEQUENCE 330 AA; 35379 MW; 7D67C5F1E5E4E915 CMC64;

Query Match 39.5%; Score 45; DB 1; Length 330;
Best Local Similarity 52.6%; Pred. No. 25;
Matches 10; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 PVPAPVTKDPATSLYTIPE 20
   1 1111 11 11
Db 107 PPPAPPTSSSATLGYGYPF 125

RESULT 11
F26_RANCA STANDARD: PRT; 470 AA.
ID F26_RANCA 091309; 091310;
AC 091309; 091310; (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 6PF-2-K/Fru-2,6-P2ASE liver/muscle Isozymes [Includes: 6-
DE phosphofructo-2-kinase (EC 2.7.1.105); Fructose-2,6-bisphosphatase
DE (EC 3.1.3.46)].
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
OX NCBI_TaxID=84400;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=liver, and muscle;
RX MEDLINE=94161720; PubMed=7509597;
RA Sakai A., Matanabe F., Furuya E.;
RT "Cloning of cDNAs for fructose 6-phosphate 2-kinase/fructose 2,6-
RT biphosphatase from frog skeletal muscle and liver, and their
RT expression in skeletal muscle.";
RL 3iochem. Biophys. Res. Commun. 198:1099-1106(1994).
CC -1- FUNCTION: SYNTHESIS AND DEGRADATION OF FRUCTOSE 2,6-BISPHOSPHATE.
CC -1- CATALYTIC ACTIVITY: ATP + D-Fructose 6-phosphate = ADP + D-
CC fructose 2,6-bisphosphate.
CC -1- CATALYTIC ACTIVITY: D-fructose 2,6-bisphosphate + H(2)O = D-
CC fructose 6-phosphate + phosphate.
CC -1- ENZYME REGULATION: PHOSPHORYLATION RESULTS IN INHIBITION OF THE
CC KINASE ACTIVITY (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

```

```

CC -----
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LIVER/L-TYPE (SHOWN HERE) AND
CC MUSCLE/M-TYPE ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO THE
CC PHOSPHOLYCERATE MUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D25223; BAA04952.1; -
DR EMBL: D25222; BAA04951.1; -
DR HSSP: P07953; 1TIP.
DR InterPro: IPR000546; 6PF2K.
DR InterPro: IPR003094; 6PF2K_kin.
DR InterPro: IPR001345; PG/BPGM_mutase.
DR Pfam: PF00300; PGAM.1.
DR Pfam: PF01591; 6PF2K.1.
DR PRINTS: PR00991; 6PF2KTNASE.
DR PRODOM: PD002665; 6PF2K.1.
DR PROSITE: PSS00175; PG_MUTASE.1.
DR Multifunctional enzyme; Transferase; Kinase; Hydrolase; ATP-binding;
DR Phosphorylation; Liver; Alternative splicing.
KW Phosphorylation; Liver; Alternative splicing.
FT DOMAIN 1 249 6-PHOSPHOFRUCTO-2-KINASE.
FT DOMAIN 250 470 6-PHOSPHOFRUCTO-2-KINASE.
FT MOD_RES 31 31 ATP (POTENTIAL).
FT NP_BIND 47 54 TO FRU-6-P (BY SIMILARITY).
FT BINDING 104 104 104 104 POTENTIAL.
FT ACT_SITE 130 130 130 130 POTENTIAL.
FT ACT_SITE 160 160 160 160 POTENTIAL.
FT BINDING 195 195 195 195 TO FRU-6-P (BY SIMILARITY).
FT ACT_SITE 258 258 258 258 FORMS THE PHOSPHOHISTIDINE INTERMEDIATE.
FT ACT_SITE 327 327 327 327 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 392 392 392 392 MADRLRELTOTRLQKIMIPHCQDLQQRGS -> MEGNRYK
FT VARSPLIC 1 31 LLEDKASRIPA (IN MUSCLE ISOFORM).
FT SEQUENCE 470 AA; 54932 MW; 53112EF3B86BD310 CMC64;

Query Match 39.5%; Score 45; DB 1; Length 470;
Best Local Similarity 40.0%; Pred. No. 36;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 PVPAPVTKDPATSLYTIPEH 21
   1 1111 11 11
Db 450 PLNVEYSRDPEDALDTVPEH 469

RESULT 12
NFTA_HUMAN STANDARD: PRT; 509 AA.
ID NFTA_HUMAN 012857; 09P2A9; 09H3X9;
AC 012857; 09P2A9; 09H3X9; (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear factor 1 A-type (Nuclear factor 1/A) (NFI-A) (NF-1/A)
DE (CCAT-box binding transcription factor) (CTF) (TGCCA-binding
DE protein).
OS Homo sapiens (Human).
GN NFTA OR KIAA1439.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";

```



```

RL DNA Res. 7:65-73(2000).
RP [2]
RN SEQUENCE OF 19-243 FROM N.A.
RX MEDLINE=96070434; PubMed=7590749;
RA Qian F., Kruse U., Lichter P., Sippel A.E.;
RT "Chromosomal localization of the four genes (NF1A, B, C, and X) for
RT the human transcription factor nuclear factor I by FISH.";
RN Genomics 28:66-73(1995).
RP [3]
RN SEQUENCE OF 235-509 FROM N.A.
RA Donnelly S.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: RECOGNIZES AND BINDS THE PALINDROMIC SEQUENCE 5'-
CC TTGCNNNNNGCCAA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN
CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE
CC INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE CTF/NF-1 FAMILY.
CC -1- SIMILARITY: CONTAINS 1 DWA/MH1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AB037860; BAA92677.1; ALT_INIT.
CC DR EMBL: U07809; AAA93124.1; -.
CC DR EMBL: AL096888; CAC10291.1; -.
CC DR TRANSFAC: T04933; -.
CC DR Genew: HGNC:7784; NF1A.
CC DR MIM: 600727;
CC DR InterPro: IPR000647; CTF_NF1_family.
CC DR InterPro: IPR003619; Dwarf1n_A.
CC DR Pfam: PF00859; CTF_NF1; 1.
CC DR SMART: SM00523; CTF_NF1; 1.
CC DR PROSITE: PS00349; CTF_NF1; 1.
CC DR Transcription regulation: DNA replication: DNA-binding; Activator;
CC KW Nuclear protein; Multigene family.
CC FT DOMAIN 67 175 DWA.
CC FT CONFLICT 186 186 A -> G (IN REF. 2).
CC FT CONFLICT 240 243 TGPN -> PAPR (IN REF. 2).
CC FT CONFLICT 509 AA: 55944 MM: 42090C6BBB229F87 CRC64;
CC SQ SEQUENCE 509 AA: 55944 MM: 42090C6BBB229F87 CRC64;

Query Match 39.5%; Score 45; DB 1; Length 509;
Best Local Similarity 57.1%; Pred. No. 39;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LPVPAPVTKDPTS 14
Db 449 VPLVPDTRKPTTS 462

RESULT 13
NF1A_CHICK STANDARD: PRT; 522 AA.
ID NF1A_CHICK STANDARD: PRT; 522 AA.
AC P17923;
DT 01-NOV-1990 (Rel. 16; Created)
DT 01-NOV-1990 (Rel. 16; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Nuclear factor 1 A-type (Nuclear factor 1(A) (NF1-A) (NF-1/A)
DE (CCAAT-box binding transcription factor) (CTF) (TGCCA-binding
DE protein).
DE GN NF1A OR NF1-A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]

```

```

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=90251434; PubMed=2339052;
RA Rupp R.A.W., Kruse U., Multhaup G., Goebel U., Beyreuther K.,
RA Sippel A.E.;
RT "Chicken NF1/TGGA proteins are encoded by at least three independent
RT genes: NF1-A, NF1-B and NF1-C with homologues in mammalian genomes.";
RN Nucleic Acids Res. 18:2607-2616(1990).
CC -1- FUNCTION: RECOGNIZES AND BINDS THE PALINDROMIC SEQUENCE 5'-
CC TTGCNNNNNGCCAA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN
CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE
CC INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF ISOFORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE CTF/NF-1 FAMILY.
CC -1- SIMILARITY: CONTAINS 1 DWA/MH1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X51486; CAA35853.1; -.
CC DR PIR: S09996; S09996.
CC DR TRANSFAC: T00544; -.
CC DR InterPro: IPR000647; CTF_NF1_family.
CC DR InterPro: IPR003619; Dwarf1n_A.
CC DR Pfam: PF00859; CTF_NF1; 1.
CC DR SMART: SM00523; CTF_NF1; 1.
CC DR PROSITE: PS00349; CTF_NF1; 1.
CC DR Transcription regulation: DNA replication: DNA-binding; Activator;
CC KW Nuclear protein; Multigene family; Alternative splicing.
CC FT DOMAIN 67 175 DWA.
CC SQ SEQUENCE 522 AA: 57473 MM: C51F3E5FE077740 CRC64;

Query Match 39.5%; Score 45; DB 1; Length 522;
Best Local Similarity 57.1%; Pred. No. 40;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LPVPAPVTKDPTS 14
Db 462 VPLVPDTRKPTTS 475

RESULT 14
CATE_ECOLI STANDARD: PRT; 753 AA.
ID CATE_ECOLI STANDARD: PRT; 753 AA.
AC P21179; P78066; P76906; P78168;
DT 01-MAY-1991 (Rel. 18; Created)
DT 01-MAY-1991 (Rel. 18; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE Catalase HPII (EC 1.11.1.6) (Hydroperoxidase II).
DE GN KATC OR B1732.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC STRAIN=K12;
RX MEDLINE=91100337; PubMed=1987146;
RA von Ossowski I., Mulvey M.R., Ieco P.A., Bortys A., Joewen P.C.;
RT "Nucleotide sequence of Escherichia coli kate, which encodes catalase
RT HPII.";
RN J. Bacteriol. 173:514-520(1991).
RN [2]
RP SEQUENCE FROM N.A.
RN STRAIN=K12 / MG1655;

```

RX MEDLINE-97426617; PubMed-9278503;  
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.R., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.,  
 RT "The complete genome sequence of *Escherichia coli* K-12";  
 RL Science 277:1453-1474(1997).  
 [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-97251357; PubMed-9097039;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Tampei G., Seki Y., Sivasubraman S., Tagami H., Takeda T.,  
 RA Takeuchi K., Takeuchi Y., Wada C., Yamamoto Y., Horinouchi T.,  
 RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map";  
 RL DNA Res. 3:363-377(1996).  
 [4]  
 RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE-95393020; PubMed-7663946;  
 RA Bravo J., Verdaguer N., Tormo J., Betzel C., Switala J., Loewen P.C.,  
 RA Fita I.,  
 RT "Crystal structure of catalase HPII from *Escherichia coli*";  
 RL Structure 3:491-502(1995).  
 [5]  
 RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE-99190072; PubMed-10091651;  
 RA Sevinc M.S., Mate M.J., Switala J., Fita I., Loewen P.C.,  
 RT "Role of the lateral channel in catalase HPII of *Escherichia coli*";  
 RL Protein Sci. 8:490-498(1999).  
 [6]  
 RN X-RAY CRYSTALLOGRAPHY (1.89 ANGSTROMS).  
 RX MEDLINE-21348730; PubMed-11455600.  
 RA Melik-Adamyan W.R., Bravo J., Carpena X., Switala J.,  
 RA Mate M.J., Fita I., Loewen P.C.,  
 RT "Substrate flow in catalases deduced from the crystal structures of  
 active site variants of HPII from *Escherichia coli*";  
 RL Proteins 44:270-281(2001).  
 CC -I- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES  
 TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.  
 CC -I- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.  
 CC -I- COFACTOR: HEME GROUP.  
 CC -I- SUBUNIT: HOMOTETRAMER.  
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -I- INDUCTION: BY ENTRY INTO STATIONARY PHASE.  
 CC -I- SIMILARITY: BELONGS TO THE CATALASE FAMILY. HPII SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M55161; AAA24039.1; -  
 DR EMBL: AE000266; AAC74802.1; -  
 DR EMBL: D90815; BAA20916.1; -  
 DR EMBL: D90816; BAA15513.1; -  
 DR EMBL: D90817; BAA15521.1; -  
 DR PIR: A39129; A39129.  
 DR PDB: 1JPH; 04-SEP-97.  
 DR PDB: 1CF9; 06-APR-99.  
 DR PDB: 1OF7; 26-APR-99.  
 DR PDB: 1G69; 30-AUG-00.  
 DR PDB: 1GGE; 30-AUG-00.  
 DR PDB: 1GGF; 30-AUG-00.  
 DR PDB: 1GGH; 30-AUG-00.  
 DR PDB: 1GGJ; 30-AUG-00.  
 DR PDB: 1GGK; 30-AUG-00.

DR EcGene: EG10509; kate.  
 DR InterPro: IPR002226; Catalase.  
 DR Pfam: PF00199; catalase; 1.  
 DR PRINTS: PR00067; CATALASE.  
 DR ProDom: PD000510; CATALASE.  
 DR PROSITE: PS00437; CATALASE\_1; 1.  
 DR PROSITE: PS00438; CATALASE\_2; 1.  
 DR Oxidoreductase; peroxidase; Iron; Heme; Hydrogen peroxide;  
 KM 3D-structure; Complete proteome.  
 FT ACT SITE 128 128 BY SIMILARITY.  
 FT ACT SITE 201 201 BY SIMILARITY.  
 FT METAL 415 415 IRON (HEME PROXIMAL LIGAND).  
 FT CONFLICT 198 198 L->F (IN REF. 2).  
 SQ SEQUENCE 753 AA; 84162 MW; 1F034E4866A70FB9 CRC64;  
 Qy 4 PAP---VTKDPATSLYTP 19  
 Db 575 PPDVNGLKRPDLSLYAIP 594  
 Best Local Similarity 39.5%; Score 45; DB 1; Length 753;  
 Matches 10; Conservative 2; Mismatches 4; Indels 4; Gaps 1;  
 RESULT 15  
 ID SRA4\_RAT STANDARD; PRT; 1048 AA.  
 AC 063627;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE CTD-binding SR-like protein Rat (fragment).  
 DE Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI-Taxid:10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hippocampus;  
 RX MEDLINE-96293459; PubMed-8692929;  
 RA Yuryev A., Patturejan M., Litingtung Y., Joshi R.V., Gentile C.,  
 RA Gebara M., Corden J.L.,  
 RT "The C-terminal domain of the largest subunit of RNA polymerase II  
 interacts with a novel set of serine/arginine-rich proteins";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:6975-6980(1996).  
 CC -I- FUNCTION: MAY ACT TO PHYSICALLY AND FUNCTIONALLY LINK  
 TRANSCRIPTION AND PRE-MRNA PROCESSING (BY SIMILARITY).  
 CC -I- SUBUNIT: INTERACT WITH THE REPETITIVE C-TERMINAL DOMAIN (CTD) OF  
 RNA POLYMERASE II (BY SIMILARITY).  
 CC -I- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U49058; AAC52660.1; -  
 DR InterPro: IPR000504; RNA\_rec\_mot.  
 DR Pfam: PF00076; rrm; 1.  
 DR SMART: SM00360; RRM; 1.  
 DR PROSITE: PS50102; RRM; 1.  
 DR PROSITE: PS00030; RRM\_RNP\_1; FALSE\_NEG.  
 KW RNA-binding.  
 FT NON\_TER 1 1  
 FT DOMAIN 3 6 POLY-PRO.  
 FT DOMAIN 155 158 POLY-PRO.  
 FT DOMAIN 422 496 RNA-BINDING (RRM).  
 FT DOMAIN 630 638 POLY-PRO.  
 FT DOMAIN 655 661 POLY-PRO.  
 FT DOMAIN 834 857 POLY-GLN.

SQ SEQUENCE 1048 AA; 113701 MW; F5CE8FD309D6883E CRC64;

Query Match 39.5%; Score 45; DB 1; Length 1048;

Best Local Similarity 72.7%; Pred. No. 84;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LPVPAPVTKDP 11  
|||||:|  
Db 583 LPVPAPITVTP 593

Search completed: May 20, 2003, 16:35:39  
Job time : 7.65385 secs

**THIS PAGE BLANK (USPTO)**



DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Putative dermal glycoprotein.  
 GN P0504E02.6  
 OS Oryza sativa (japonica cultivar-group)  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriophytaceae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0504E02."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003269; BAB9707.1;  
 SO SEQUENCE 424 AA; 44620 MW; 0DACD5C852C34 CRC64;

Query Match 57.9%; Score 66; DB 10; Length 424;  
 Best Local Similarity 66.7%; Pred. No. 0.15;  
 Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 PVPAPVTKDPATSLYTI 19  
 DB 34 PLVAATTKDAATSLYTPV 51

RESULT 3  
 ID 08XUH5 PRELIMINARY; PRT; 173 AA.  
 AC 08XUH5;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Hypothetical protein RSC3213.  
 GN RSC3213 OR RS02430.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Brotlier P., Camus J.C., Catolico L.,  
 RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Sautin W., Schlex T.,  
 RA Sigler P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 DR EMBL: AL646074; CAD17001.1;  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 173 AA; 1865 MW; F902A56F9D25E737 CRC64;

Query Match 46.5%; Score 53; DB 16; Length 173;  
 Best Local Similarity 64.3%; Pred. No. 4.7;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PVPAPVTKDPATSL 15  
 DB 72 PVPAPVTKDPATSL 85  
 RESULT 4  
 ID 08S1V3 PRELIMINARY; PRT; 422 AA.  
 AC 08S1V3;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Putative dermal glycoprotein.

GN P0504E02.6  
 OS Oryza sativa (japonica cultivar-group)  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriophytaceae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0504E02."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003269; BAB9705.1;  
 SO SEQUENCE 422 AA; 44745 MW; 08CFE26346400A9F CRC64;

Query Match 45.6%; Score 52; DB 10; Length 422;  
 Best Local Similarity 64.7%; Pred. No. 16;  
 Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 PVPAPVTKDPATSLYTI 18  
 DB 37 PLVTAATTKDGAATKLYTI 53

RESULT 5  
 ID 08QVD2 PRELIMINARY; PRT; 492 AA.  
 AC 08QVD2;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE RNA-dependent RNA polymerase (Fragment).  
 OS Colorado tick fever virus.  
 OC Viruses; dsRNA viruses; Reoviridae; Coltivirus.  
 OX NCBI\_TaxID=46839;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CTFV-S6-14-03;  
 RX MEDLINE=21954629; PubMed=11958454;  
 RA Attoui H., Mohd J.F., Biagini P., Cantaloube J.F., de M.P.,  
 RA Murphy F.A., de L.X.;  
 RT "Genus Coltivirus (family Reoviridae): genomic and morphologic  
 characterization of Old World viruses";  
 RL Arch. Virol. 147:533-561(2002).  
 DR EMBL: AF343051; AAM18357.1;  
 KW RNA-directed RNA polymerase.  
 FT NON\_TER 1  
 FT NON\_TER 492  
 FT NON\_TER 1  
 SO SEQUENCE 492 AA; 55966 MW; 501250E88F84DA58 CRC64;

Query Match 45.6%; Score 52; DB 12; Length 492;  
 Best Local Similarity 52.2%; Pred. No. 19;  
 Matches 12; Conservative 2; Mismatches 5; Indels 4; Gaps 1;

QY 1 LPVPAPVTKD-----PATSLYTI 19  
 DB 408 VPPEPLVOKDELICPGTSVYTI 430  
 RESULT 6  
 ID 099I35 PRELIMINARY; PRT; 991 AA.  
 AC 099I35;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE RNA polymerase (Fragment).  
 OS Colorado tick fever virus.  
 OC Viruses; dsRNA viruses; Reoviridae; Coltivirus.  
 OX NCBI\_TaxID=46839;  
 RN [1]  
 RP SEQUENCE FROM N.A.







ID 096SL8 PRELIMINARY; PRT; 496 AA.  
 AC 096SL8;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE CDNA FLJ14768 fis, clone NTZRP3004125, moderately similar to Mus musculus zinc finger protein splice variant FIZ1-B (FIZ1) mRNA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Wadatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Chiba Y., Ishida S., Murakami K., Ono Y., Takiguchi S., Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Ninomiya K., Iwayanagi T.,  
 RA "NEDO human cDNA sequencing project."  
 RT Submitted (May-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AK027674; BAB55286.1;  
 DR InterPro: IPR000564; 2FE2S\_ferredoxin.  
 DR InterPro: IPR000822; ZnF\_C2H2.  
 DR Pfam: PF00096; ZnF\_C2H2; 1.  
 DR ProDom: PD000003; ZnF\_C2H2; 1.  
 DR PROSITE: PS00197; 2FE2S\_FERREDOXIN; UNKNOWN\_1.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_10.  
 DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 1.  
 DR DNA-binding: Zinc-finger.  
 SQ SEQUENCE 496 AA; 51993 MW; 937EEC713FFBD58F CRC64;

Query Match 43.9%; Score 50; DB 4; Length 496;  
 Best Local Similarity 38.1%; Pred. No. 37;  
 Matches 8; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 LPVPAVTRKPAATSLYTIPTF 21  
 :| | | | | : |||  
 DB 4 VPAPTRPAPAPAAAPRVDPFH 24

RESULT 15  
 09PVZ2 PRELIMINARY; PRT; 2037 AA.  
 ID 09PVZ2;  
 AC 09PVZ2;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Nucleoporin CAN.  
 DE Nucleoporin CAN.  
 GN CAN.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxId=8335;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-99384295; PubMed-10454574;  
 RA Askjaer P., Bach A., Wilm M., Bischoff R., Weeks D.L., Ogniewski V., Ohno M., Niehrs C., Kjems J., Matla J.W., Fornerod M.,  
 RT "RangRP-regulated interactions of CRM1 with nucleoporins and a shuttling DEAD-box helicase."  
 RL Mol. Cell. Biol. 19:6276-6285(1999).  
 DR EMBL: AJ243889; CAB53357.1; -;  
 DR InterPro: IPR000515; BPD\_transp.  
 DR InterPro: IPR004325; Nucleoporin\_FG.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF03093; Nucleoporin\_FG; 16.  
 DR Pfam: PF00400; WD40; 2.  
 DR SMART: SM00320; WD40; 2.  
 DR PROSITE: PS00402; BPD\_TRANSF\_INN\_MEMBR; UNKNOWN\_1.  
 KW Porin; Repeat; WD repeat.

SQ SEQUENCE 2037 AA; 208932 MW; 2250EA49A0435635 CRC64;  
 Query Match 43.9%; Score 50; DB 13; Length 2037;  
 Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 5 APYTKDPATSLYTIPTF 20  
 :| | | | | : |||  
 DB 1672 APLMKGPATSGSTRIPF 1687

Search completed: May 20, 2003, 16:43:13  
 Job time : 28.9615 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.P5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:35:02 ; Search time 9.69231 Seconds  
(without alignments)  
63,750 Million cell updates/sec

Title: US-09-869-155-15

Sequence: 1 LPVPAPYTKDPATSLYRPFH 21

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	39.5	690	4	US-08-971-089-6
2	45	39.5	822	4	US-09-564-805-222
3	44	38.6	353	1	US-08-073-807A-17
4	44	38.6	380	1	US-08-073-807A-18
5	44	38.6	416	1	US-08-073-807A-2
6	43.5	38.2	978	2	US-08-415-593-43
7	43	37.7	288	4	US-09-651-200-14
8	43	37.7	292	4	US-09-651-200-16
9	43	37.7	297	2	US-08-580-545B-6
10	43	37.7	297	4	US-09-262-653A-6
11	43	37.7	322	2	US-08-892-880-2
12	43	37.7	322	4	US-09-232-160-21
13	43	37.7	751	4	US-09-036-987A-24
14	43	37.7	751	4	US-09-370-700-24
15	42.5	37.3	432	2	US-08-749-902-6
16	42	36.8	135	4	US-09-134-001C-3069
17	42	36.8	383	4	US-09-134-001C-3068
18	42	36.8	693	1	US-08-553-279-2
19	42	36.8	826	4	US-09-564-805-2
20	42	36.8	826	4	US-09-564-805-224
21	42	36.8	826	4	US-09-564-805-226
22	42	36.8	846	4	US-09-134-001C-3067
23	42	36.8	1018	1	US-08-072-610-2
24	42	36.8	1018	2	US-08-719-822B-2
25	42	36.8	1018	4	US-09-092-458-2
26	42	36.8	1183	4	US-09-134-001C-3530
27	42	36.8	2182	2	US-08-487-826B-16

28	41.5	36.4	29	2	US-08-945-168-103	Sequence 103, Ap
29	41.5	36.4	379	3	US-09-320-878-6	Sequence 6, Appl
30	41.5	36.4	379	4	US-09-105-537-16	Sequence 16, Appl
31	41.5	36.4	488	1	US-07-794-393-2	Sequence 2, Appl
32	41.5	36.4	488	1	US-08-001-711-2	Sequence 2, Appl
33	41.5	36.4	488	3	US-08-704-711A-22	Sequence 22, Appl
34	41.5	36.4	488	4	US-09-521-220-22	Sequence 22, Appl
35	41.5	36.4	488	4	US-09-391-104-31	Sequence 31, Appl
36	41.5	36.4	489	4	US-08-448-489-11	Sequence 11, Appl
37	41.5	36.4	3782	4	US-09-105-537-4	Sequence 4, Appl
38	41	36.0	112	4	US-09-134-001C-5057	Sequence 5057, Ap
39	41	36.0	178	1	US-08-928-443-1	Sequence 1, Appl
40	41	36.0	178	3	US-09-129-055-1	Sequence 1, Appl
41	41	36.0	279	1	US-08-872-979-5	Sequence 5, Appl
42	41	36.0	289	1	US-08-393-985-8	Sequence 8, Appl
43	41	36.0	289	3	US-08-872-979-6	Sequence 6, Appl
44	41	36.0	310	3	US-08-796-792-4	Sequence 4, Appl
45	41	36.0	480	4	US-09-108-020-6	Sequence 6, Appl

## ALIGNMENTS

RESULT 1  
US-08-971-089-6  
; Sequence 6, Application US/08971089  
; Patent No. 6376174  
; GENERAL INFORMATION:  
; APPLICANT: Pulst, Stefan M.  
; APPLICANT: Scoles, Daniel R.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING  
; TITLE OF INVENTION: SCHWANNOMIN-BINDING-PROTEINS AND PRODUCTS RELATED THERETO  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CAMPBELL & FLORES, LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/971,089  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/030,987  
; FILING DATE: 15-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ramos, Robert T.  
; REGISTRATION NUMBER: 37,915  
; REFERENCE/DOCKET NUMBER: P-CE 2862  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619)535-9001  
; TELEFAX: (619)535-8949  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 690 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-971-089-6  
Query Match 39.5%; Score 45; DB 4; Length 690;  
Best Local Similarity 44.4%; Pred. No. 69;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
QY 2 PVPAPYTKDPATSLYRPF 19  
|:|:|:|:|:|

Db 294 PMPSSASAPPASSLYSP 311

## RESULT 2

US-09-564-805-222  
Sequence 222, Application US/09564805

Patent No. 6333403

GENERAL INFORMATION:

APPLICANT: Tavligian, Sean V.

APPLICANT: Teng, David H.F.

APPLICANT: Stmard, Jacques

APPLICANT: Rommens, Johanna M.

APPLICANT: Myriad Genetics, Inc.

TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility

TITLE OF INVENTION: Gene and a Paralog and Orthologous Genes

FILE REFERENCE: 2318-258

CURRENT APPLICATION NUMBER: US/09/564,805

CURRENT FILING DATE: 2000-05-05

PRIOR FILING DATE: 1998-11-06

PRIOR APPLICATION NUMBER: 09/434,382

PRIOR FILING DATE: 1999-11-05

NUMBER OF SEQ ID NOS: 240

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO: 222

LENGTH: 822

TYPE: PRT

ORGANISM: Mus musculus

US-09-564-805-222

Query Match: 39.5%; Score 45; DB 4; Length 822;

Best Local Similarity 40.0%; Pred. No. 86;

Matches 8; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Db 159 PHSAPRYKDEMTYVOPH 178

## RESULT 3

US-08-073-807A-17  
Sequence 17, Application US/08073807A

Patent No. 5646248

GENERAL INFORMATION:

APPLICANT: Sawada, Ritsuko

APPLICANT: Lowe, John B.

APPLICANT: Fukuda, Minoru

TITLE OF INVENTION: CELL SURFACE LAMP EXPRESSION AND

TITLE OF INVENTION: SELECTIN-DEPENDENT ADHESION

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: United States

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/073,807A

FILING DATE: 08-JUN-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 9567

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 353 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-073-807A-17

Query Match 38.6%; Score 44; DB 1; Length 353;

Best Local Similarity 41.2%; Pred. No. 44;

Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 180 PPSPPVKSPSPVDKYNV 196

## RESULT 4

US-08-073-807A-18  
Sequence 18, Application US/08073807A

Patent No. 5646248

GENERAL INFORMATION:

APPLICANT: Sawada, Ritsuko

APPLICANT: Lowe, John B.

APPLICANT: Fukuda, Minoru

TITLE OF INVENTION: CELL SURFACE LAMP EXPRESSION AND

TITLE OF INVENTION: SELECTIN-DEPENDENT ADHESION

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: United States

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/073,807A

FILING DATE: 08-JUN-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 9567

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 380 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-073-807A-18

Query Match 38.6%; Score 44; DB 1; Length 380;

Best Local Similarity 41.2%; Pred. No. 48;

Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 207 PPSPPVKSPSPVDKYNV 223

## RESULT 5

US-08-073-807A-2  
Sequence 2, Application US/08073807A

Patent No. 5646248

GENERAL INFORMATION:

APPLICANT: Sawada, Ritsuko

APPLICANT: Lowe, John B.

APPLICANT: Fukuda, CELU  
TITLE OF INVENTION: CELL SURFACE LAMP EXPRESSION AND  
TITLE OF INVENTION: SELECTIN-DEPENDENT ADHESION  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
City: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/073,807A  
FILING DATE: 08-JUN-1993  
CLASSIFICATION: .435  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 9567  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 416 amino acids  
type: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-073-807A-2

Query Match	38.6%	Score 44	DB 1	Length 416
Best Local	Similarity 41.2%	Pred. No. 53		
Matches 7	Conservative 3	Mismatches 7	Indels 0	Gaps 0

```
QY      2 PVPAPVTKDPATSLYTI 18
        | | : | | | : | :
db      207 PSPSPVPKSPSPVDKYNV 223
```

RESULT 6  
 US-08-415-593-43  
 Sequence 43, Application US/08415593  
 Patent No. 5912140  
 Patent No. 5912140 5776726  
 GENERAL INFORMATION:  
 APPLICANT: Whoriskey, Susan K.  
 APPLICANT: Quinn, Cheryl L.  
 APPLICANT: Tao, Nianjun  
 APPLICANT: Pollits-Virk, Karen I.  
 APPLICANT: Schimmel, Paul R.  
 TITLE OF INVENTION: Recombinant Pneumocystis Aminoacyl tRNA  
 NUMBER OF INVENTION: Synthetase Genes, Tester Strains and Assays  
 NUMBER OF SEQUENCES: 46  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 STREET: Two Millita Drive  
 CITY: Lexington  
 STATE: Massachusetts  
 COUNTRY: U.S.  
 ZIP: 02173  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/415,593  
 FILING DATE:

```

1 CLASSIFICATION: 4355
2
3 ATTORNEY/AGENT INFORMATION:
4
5 NAME: BROOK, David E.
6
7 REGISTRATION NUMBER: 22,592
8
9 REFERENCE/DOCKET NUMBER: CPI94-09
10
11 TELECOMMUNICATION INFORMATION:
12
13 TELEPHONE: 617-861-6240
14
15 TELEFAX: 617-862-9540
16
17 INFORMATION FOR SEQ ID NO: 43:
18
19 SEQUENCE CHARACTERISTICS:
20
21 LENGTH: 978 amino acids
22
23 TYPE: amino acid
24
25 TOPOLOGY: linear
26
27 MOLECULE TYPE: peptide
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
10
```

Query Match	38.2%	Score 43.5;	DB 2;	Length 978;
Best Local Similarity	39.1%;	Pred. No. 1.8e+02;		
Matches	9;	Conservative	5;	Mismatches 2;
			Indels	7;
			Gaps	1;

```
QY      6 PVTKDPATSL-----YTIPEH 21
          | : : | : | | | : | : | |
Db     116 PLEDPSTSLAWTTTPWTLP SH 138
```

```

RESULT 7
US-09-651-200-14
; Sequence 14, Application US/09651200
; Patent No. 6429303
; GENERAL INFORMATION:
; APPLICANT: Green et al
; TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
; TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
; TITLE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: 15966-562 (CUBA-62)
; CURRENT APPLICATION NUMBER: US/09/651,200
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/152383
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/172909
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/183578
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Macaca mulatta
US-09-651-200-14

```

Query Match	37.7%;	Score 43;	DB 4;	Length 288;
Best Local Similarity	43.8%;	Pred. No. 48;		
Matches	7;	Conservative	4;	Mismatches 5;
				Indels 0;
				Gaps 0;

```
QY      3 VPAPVTKDPATSLYTI 18
      : |::|| | |||:
Db     185 ISTVVSQDPETELTYV 200
```

```

? RESULT 8
? US-09-651-200-16
? Sequence 16, Application US/09651200
? Patent No. 6429303
? GENERAL INFORMATION:
? APPLICANT: Green et al
? TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
? TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
? TITLE OF INVENTION: Polypeptides Encoded Thereby
? FILE REFERENCE: 15966-562 (CWA-62)
? CURRENT APPLICATION NUMBER: US/09/651,200
? CURRENT FILING DATE: 2000-08-30
? PRIOR APPLICATION NUMBER: 60/152363
?

```



```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS-
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/892,880
; FILING DATE: HEREWITH
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/021,762
; FILING DATE: 15-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488, 0490001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 322 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-892-880-2

Query Match          37.7%; Score 43; DB 2; Length 322;
Best Local Similarity 47.1%; Pred. No. 55;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 VPAPVTKDPATSLYTIP 19
:| | | | | :| |
Db 178 IPAPTTTPPAPASTIP 194

RESULT 12
US-09-232-160-21
; Sequence 21, Application US/09232160
; Patent No. 6368794
; GENERAL INFORMATION:
; APPLICANT: Steve Daniel
; APPLICANT: James Gilmore
; APPLICANT: Susan G. Stuart
; APPLICANT: Laura Stuve
; TITLE OF INVENTION: DETECTION OF ALTERED EXPRESSION OF GENES REGULATING CELL
; FILE REFERENCE: PA-0003 US
; CURRENT APPLICATION NUMBER: US/09/232,160
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PERL Program
; SEQ ID NO 21
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 3044710
US-09-232-160-21

Query Match          37.7%; Score 43; DB 4; Length 322;
Best Local Similarity 47.1%; Pred. No. 55;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 VPAPVTKDPATSLYTIP 19
:| | | | | :| |
Db 178 IPAPTTTPPAPASTIP 194

RESULT 13
US-09-036-987A-24
; Sequence 24, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
```

```

;
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Walidon, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R.
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 751 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-09-036-987A-24

Query Match          37.7%; Score 43; DB 4; Length 751;
Best Local Similarity 31.6%; Pred. No. 16+02;
Matches 6; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 3 VPAPVTKDPATSLYTIP 21
:| :| :| :| :| :|
Db 282 IPTDDEVPVALVLPFH 300

RESULT 14
US-09-370-700-24
; Sequence 24, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Walidon, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 24
; LENGTH: 751
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
```

US-09-370-700-24

Query Match 37.7%; Score 43; DB 4; Length 751;

Best Local Similarity 31.6%; Pred. No. 1.6e+02;  
Matches 6; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 3 VPAPVTKDPATSLVTYPE 21

DB 282 IPTDDDEVPTVAIYLVPFH 300

RESULT 15

US-08-749-902-6

Sequence 6, Application US/08749902  
Patent No. 5985635

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Goll, Surya K.

APPLICANT: Hillman, Jennifer L.

TITLE OF INVENTION: NOVEL HUMAN SERINE/THREONINE

TITLE OF INVENTION: PROTEIN KINASES

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSEO Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/749,902

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0150 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 432 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 1016651

US-08-749-902-6

Query Match 37.3%; Score 42.5; DB 2; Length 432;

Best Local Similarity 42.9%; Pred. No. 94;

Matches 9; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

QY 1 LPV-PAPVTKDPATSLVTYPE 20

DB 323 VPFPSPETKDRRWRLTVVY 343

Search completed: May 20, 2003, 16:44:40  
Job time: 11.6923 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:49:49 ; Search time 10.3846 Seconds  
(without alignments)  
200.549 Million cell updates/sec

Title: US-09-869-155-15

Sequence: 1 LPVPAPYTKDPATSLYIPFH 21

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/PCIT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/PCITUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	42.1	81	10 US-09-864-761-47685	Sequence 47685, A
2	47	41.2	750	10 US-09-815-242-13985	Sequence 13985, A
3	46	40.4	2701	9 US-10-171-311-83	Sequence 83, Appl
4	45	39.5	136	9 US-10-091-483-211	Sequence 211, App
5	45	39.5	136	10 US-09-764-846-211	Sequence 211, App
6	45	39.5	635	9 US-09-738-626-6614	Sequence 6614, App
7	45	39.5	690	9 US-10-117-604-6	Sequence 6, Appl
8	45	39.5	753	10 US-09-815-242-10181	Sequence 10181, A
9	45	39.5	822	9 US-09-988-626-222	Sequence 222, App
10	45	39.5	822	9 US-09-988-667-222	Sequence 222, App
11	45	39.5	1708	9 US-09-468-147-166	Sequence 166, App
12	44.5	39.0	532	9 US-10-225-928-2	Sequence 2, Appl
13	44	38.6	475	10 US-09-927-091-1	Sequence 1, Appl
14	44	38.6	504	8 US-08-980-068B-21	Sequence 21, Appl
15	43.5	38.2	328	9 US-10-243-739-11	Sequence 11, Appl
16	43.5	38.2	375	12 US-10-062-254-373	Sequence 373, App
17	43.5	38.2	580	10 US-09-808-387-36	Sequence 36, Appl
18	43	37.7	138	10 US-09-925-301-1164	Sequence 1164, App
19	43	37.7	156	9 US-09-907-969-582	Sequence 582, App

20	43	37.7	292	10 US-09-303-510-2	Sequence 2, Appl
21	43	37.7	292	10 US-09-303-510-4	Sequence 4, Appl
22	43	37.7	292	10 US-09-303-040-2	Sequence 2, Appl
23	43	37.7	292	10 US-09-303-040-4	Sequence 4, Appl
24	43	37.7	294	9 US-09-907-969-487	Sequence 487, App
25	43	37.7	294	10 US-09-884-441-487	Sequence 487, App
26	43	37.7	322	9 US-09-905-291A-201	Sequence 201, App
27	43	37.7	322	9 US-10-063-547-6	Sequence 6, Appl
28	43	37.7	322	9 US-09-902-853-201	Sequence 201, App
29	43	37.7	322	9 US-09-907-824-201	Sequence 201, App
30	43	37.7	322	9 US-09-907-841-201	Sequence 201, App
31	43	37.7	322	9 US-09-904-011-201	Sequence 201, App
32	43	37.7	322	9 US-10-063-616-6	Sequence 6, Appl
33	43	37.7	322	9 US-09-906-742-201	Sequence 201, App
34	43	37.7	322	9 US-10-063-502-6	Sequence 6, Appl
35	43	37.7	322	9 US-09-906-838-201	Sequence 201, App
36	43	37.7	322	9 US-09-907-613-201	Sequence 201, App
37	43	37.7	322	9 US-09-907-942-201	Sequence 201, App
38	43	37.7	322	9 US-09-904-820-201	Sequence 201, App
39	43	37.7	322	9 US-09-904-859-201	Sequence 201, App
40	43	37.7	322	9 US-09-909-204-201	Sequence 201, App
41	43	37.7	322	9 US-09-904-786-201	Sequence 201, App
42	43	37.7	322	9 US-09-906-646-201	Sequence 201, App
43	43	37.7	322	9 US-09-906-700-201	Sequence 201, App
44	43	37.7	322	9 US-09-902-903-201	Sequence 201, App
45	43	37.7	322	9 US-09-903-749A-201	Sequence 201, App

#### ALIGNMENTS

##### RESULT 1

US-09-864-761-47685  
Sequence 47685, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecmca-x-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263,6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 47685  
LENGTH: 81  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC009305.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.57  
OTHER INFORMATION: EST\_HUMAN HIT: BE881567.1, EVALUO 3.00e-04  
OTHER INFORMATION: SWISSPROT HIT: Q09472, EVALUO 2.00e-04  
US-09-364-761-47685

Query Match 42.1%; Score 48; DB 10; Length 81;  
Best Local Similarity 45.0%; Pred. No. 7.8;  
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY .1 LPVPAPVTKDPATSLVYTF 20  
||: ||: ||: ||: ||: ||  
Db 48 LPATVPVQPPQTSIDSVTF 67

RESULT 2  
US-09-815-242-13985  
Sequence 13985, Application US/09815242  
Patent No. US20020061569A1

GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13985  
LENGTH: 750  
TYPE: PRT  
ORGANISM: Salmonella typhi

US-09-815-242-13985  
Query Match 41.2%; Score 47; DB 10; Length 750;  
Best Local Similarity 50.0%; Pred. No. 1.4e+02;  
Matches 10; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

QY 4 PAP---VTKDPATSLVYTF 19  
||: ||: ||: ||: ||: ||  
Db 572 PPPDVNGLKDPALSLVAVP 591

RESULT 3  
US-10-171-311-83  
Sequence 83, Application US/10171311  
Publication No. US20030087270A1

GENERAL INFORMATION:  
APPLICANT: Schlegel, Robert  
APPLICANT: Chen, Yan  
APPLICANT: Zhao, Xumei  
APPLICANT: Monahan, John  
APPLICANT: Kamatkar, Shubhangi  
APPLICANT: Gattai, Karen  
APPLICANT: Gannavarapu, Manjula  
APPLICANT: Hoerish, Sebastian  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
FILE REFERENCE: MRI-035  
CURRENT APPLICATION NUMBER: US/10/171,311  
CURRENT FILING DATE: 2002-06-12  
PRIOR APPLICATION NUMBER: US 60/298,159  
PRIOR FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/298,155  
PRIOR FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/335,936  
PRIOR FILING DATE: 2001-11-14  
NUMBER OF SEQ ID NOS: 238  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 83  
LENGTH: 2701  
TYPE: PRT  
ORGANISM: Homo sapiens

US-10-171-311-83  
Query Match 40.4%; Score 46; DB 9; Length 2701;  
Best Local Similarity 44.4%; Pred. No. 8.1e+02;  
Matches 8; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 2 PVPAPVTKDPATSLVYTF 19  
||: ||: ||: ||: ||: ||  
Db 1874 PTPAPASSPAPVITAP 1891

RESULT 4  
US-10-091-483-211  
Sequence 211, Application US/10091483  
Publication No. US20030049650A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT212C1  
CURRENT APPLICATION NUMBER: US/10/091,483  
CURRENT FILING DATE: 2002-03-07  
NUMBER OF SEQ ID NOS: 348  
Prior Application removed - See File Wrapper or Palm  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 211  
LENGTH: 136  
TYPE: PRT  
ORGANISM: Homo sapiens

US-10-091-483-211  
Query Match 39.5%; Score 45; DB 9; Length 136;  
Best Local Similarity 57.1%; Pred. No. 37;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LPVPAPVTKDPATSLVYTF 14  
||: ||: ||: ||: ||: ||

Db 76 VPLPVPTKPPPTS 89

RESULT 5  
US-09-764-846-211

; Sequence 211, Application US/09764846  
; Patent No. US20020102638A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT212

; CURRENT APPLICATION NUMBER: US/09/764,846

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 348

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO: 211

; LENGTH: 136

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-764-846-211

Query Match 39.5%; Score 45; DB 10; Length 136;  
Best Local Similarity 57.1%; Pred. No. 37;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LPVPAPVTKDPATSP 14

Db 76 VPLPVPTKPPPTS 89

RESULT 6  
US-09-738-626-6614

; Sequence 6614, Application US/09738626  
; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAMA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAOKO

; APPLICANT: SENO, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: PatentIn ver. 3.0

; SEQ ID NO: 6614

; LENGTH: 635

; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum

US-09-738-626-6614

Query Match 39.5%; Score 45; DB 9; Length 635;  
Best Local Similarity 50.0%; Pred. No. 2.1e+02;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 PVPAPVTKDPATSLYTP 19

Db 181 PVPAPVTSAPGASVPSIP 198

RESULT 7  
US-10-117-604-6

; Sequence 6, Application US/10117604  
; Patent No. US20020168672A1

; GENERAL INFORMATION:

; APPLICANT: Pulst, Stefan M.

; APPLICANT: Sciles, Daniel R.

; TITLE OF INVENTION: NUCLEIC ACID ENCODING

; SCHWANNOMIN-BINDING-PROTEINS AND PRODUCTS RELATED THEREIN

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESSES:

; ADDRESS: CAMPBELL & FLORES, LLP

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/117,604

; FILING DATE: 04-Apr-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/971,089

; FILING DATE: <Unknown>

; APPLICATION NUMBER: US 60/030,987

; FILING DATE: 15-Nov-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Ramos, Robert T.

; REGISTRATION NUMBER: 37,915

; REFERENCE/DOCKET NUMBER: P-CE 2862

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619)535-9001

; TELEFAX: (619)535-8949

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 690 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-10-117-604-6

Query Match 39.5%; Score 45; DB 9; Length 690;  
Best Local Similarity 44.4%; Pred. No. 2.4e+02;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 PVPAPVTKDPATSLYTP 19

Db 294 PMSASSAPPASSLYSSP 311

RESULT 8  
US-09-815-242-10181

; Sequence 10181, Application US/09815242  
; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

```

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10181
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10181

```

```

Query Match      39.5%; Score 45; DB 9; Length 753;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 10; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

```

```

QY 4 PAP---VTKDPATSLYTIPT 19
    | | | | | | | | | | | | |
DB 575 PPPDVNGLKDPDSLTYALT 594

```

```

RESULT 9
US-09-988-626-222
; Sequence 222, Application US/09988626
; Publication No. US20030044959A1
; GENERAL INFORMATION:
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.F.
; APPLICANT: Simard, Jacques
; APPLICANT: Rommens, Johanna M.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility
; FILE REFERENCE: 2318-258
; CURRENT APPLICATION NUMBER: US/09/988,626
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 09/564,805
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/107,468
; PRIOR FILING DATE: 1998-11-06
; PRIOR APPLICATION NUMBER: 09/434,382
; PRIOR FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 222
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-988-626-222

```

```

Query Match      39.5%; Score 45; DB 9; Length 822;
Best Local Similarity 40.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 2 PVPAPYTKDPATSLYTIPT 21
    | | | | | | | | | | | | |
DB 159 PHSAPYKDEMTYVQVPIH 178

```

```

RESULT 10
US-09-988-687-222
; Sequence 222, Application US/09988687

```

```

; Publication No. US20030045704A1
; GENERAL INFORMATION:
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.F.
; APPLICANT: Simard, Jacques
; APPLICANT: Rommens, Johanna M.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility
; FILE REFERENCE: 2318-258
; CURRENT APPLICATION NUMBER: US/09/988,687
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 09/564,805
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/107,468
; PRIOR FILING DATE: 1998-11-06
; PRIOR APPLICATION NUMBER: 09/434,382
; PRIOR FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 222
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-988-687-222

```

```

Query Match      39.5%; Score 45; DB 9; Length 822;
Best Local Similarity 40.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 2 PVPAPYTKDPATSLYTIPT 21
    | | | | | | | | | | | | |
DB 159 PHSAPYKDEMTYVQVPIH 178

```

```

RESULT 11
US-09-468-147-166
; Sequence 166, Application US/09468147A
; Publication No. US20030049601A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Schlauder, George G.
; APPLICANT: Etker, James C.
; APPLICANT: Desai, Suresh M.
; APPLICANT: Dawson, George J.
; APPLICANT: Mushahwar, I. K.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
; FILE REFERENCE: 6232-US-P1
; CURRENT APPLICATION NUMBER: US/09/468,147A
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: US 09/173,141
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: US 60/061,199
; EARLIER FILING DATE: 1997-10-15
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 166
; LENGTH: 1708
; TYPE: PRT
; ORGANISM: Hepatitis E Virus
; FEATURE:
; OTHER INFORMATION: Xaa = Unknown or Other at position 322
; OTHER INFORMATION: Xaa = Unknown or Other at position 331
; OTHER INFORMATION: Xaa = Unknown or Other at position 445
; OTHER INFORMATION: Xaa = Unknown or Other at position 448
; OTHER INFORMATION: Xaa = Unknown or Other at position 634
; OTHER INFORMATION: Xaa = Unknown or Other at position 646
; OTHER INFORMATION: Xaa = Unknown or Other at position 811
; OTHER INFORMATION: Xaa = Unknown or Other at position 1533
; OTHER INFORMATION: Xaa = Unknown or Other at position 1578
; OTHER INFORMATION: Xaa = Unknown or Other at position 1691
US-09-468-147-166

```

Query Match 39.5%; Score 45; DB 9; Length 1708;  
Best Local Similarity 42.3%; Pred. No. 6.6e+02;  
Matches 11; Conservative 1; Mismatches 6; Indels 8; Gaps 1;

OY 2 PVPAPVTKDPATSLYTIPTF 19  
Db 775 PPPPPVRRKPPATSPPTRRRLTYTP 800

RESULT 12  
US-10-225-928-2  
Sequence 2, Application US/10225928  
Publication No. US20030092041A1  
GENERAL INFORMATION:  
APPLICANT: White, David  
TITLE OF INVENTION: NOVEL USE FOR MUSCARINIC RECEPTOR M5 IN  
FILE REFERENCE: MP101-210P1RM  
CURRENT APPLICATION NUMBER: US/10/225,928  
CURRENT FILING DATE: 2002-08-22  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 532  
TYPE: PRT  
ORGANISM: human  
US-10-225-928-2

Query Match 39.0%; Score 44.5; DB 9; Length 532;  
Best Local Similarity 45.5%; Pred. No. 2.1e+02;  
Matches 10; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

OY 1 LPVPAPVTKDPATSLYTIPTF 21  
Db 403 MCPPPVAKEPSTKGLNPNPSH 424

RESULT 13  
US-09-927-091-1  
Sequence 1, Application US/09927091  
Patent No. US20020119541A1  
GENERAL INFORMATION:  
APPLICANT: KILLARY, ANN  
APPLICANT: LOFT, STEVE  
APPLICANT: CHANDLER, DAWN  
TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1  
FILE REFERENCE: UTSC:6510S  
CURRENT APPLICATION NUMBER: US/09/927,091  
CURRENT FILING DATE: 2001-08-09  
PRIOR APPLICATION NUMBER: 60/227,560  
PRIOR FILING DATE: 2000-08-23  
PRIOR APPLICATION NUMBER: 60/225,033  
PRIOR FILING DATE: 2000-08-10  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 475  
TYPE: PRT  
ORGANISM: Human  
US-09-927-091-1

Query Match 38.6%; Score 44; DB 10; Length 475;  
Best Local Similarity 61.5%; Pred. No. 2.1e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 PVPAPVTKDPATSLYTIPTF 14  
Db 292 PVPALTLDPGTA 304

RESULT 14  
US-08-980-068B-21

Sequence 21, Application US/08980068B  
Patent No. US20020081718A1  
GENERAL INFORMATION:  
APPLICANT: HOTTA, Yoshiaki  
TITLE OF INVENTION: A POLYPEPTIDE COMMON TO GLIAL CELLS MISSING (GCM)  
FILE REFERENCE: 97-1513\*/LC/00653  
CURRENT APPLICATION NUMBER: US/08/980,068B  
CURRENT FILING DATE: 1997-11-26  
NUMBER OF SEQ ID NOS: 23  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 21  
LENGTH: 504  
TYPE: PRT  
ORGANISM: Drosophila  
US-08-980-068B-21

Query Match 38.6%; Score 44; DB 8; Length 504;  
Best Local Similarity 61.5%; Pred. No. 2.3e+02;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 LPVPAPVTKDPATSLYTIPTF 13  
Db 14 MPVPMVPSPAT 26

RESULT 15  
US-10-243-739-11  
Sequence 11, Application US/10243739  
Publication No. US20030091593A1  
GENERAL INFORMATION:  
APPLICANT: Bachmann, Martin F.  
APPLICANT: Storni, Tazio  
APPLICANT: Lechner, Franziska  
TITLE OF INVENTION: In vivo Activation of Antigen Presenting Cells for Enhancement  
FILE REFERENCE: 1700.0210001  
CURRENT APPLICATION NUMBER: US/10/243,739  
CURRENT FILING DATE: 2002-09-16  
PRIOR APPLICATION NUMBER: 60/318,967  
PRIOR FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 11  
LENGTH: 328  
TYPE: PRT  
ORGANISM: Bacteriophage Q-beta  
US-10-243-739-11

Query Match 38.2%; Score 43.5; DB 9; Length 328;  
Best Local Similarity 45.0%; Pred. No. 1.6e+02;  
Matches 9; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

OY 1 LPVPAPVTKDPATSLYTIPTF 20  
Db 150 IPDP-PIDPPCTGKVTCTPF 168

Search completed: May 20, 2003, 18:00:01  
Job time: 11.3846 secs

**THIS PAGE BLANK (USPTO)**



XX Claim 24; Page 112; 112pp; English.  
PS  
CC The present sequence is derived from an endo-beta-1,4-xylanase  
CC inhibitor. The protein is obtained from wheat flour. The specification  
CC also describes a mutant xylanase protein. The xylanase is useful for  
CC preparing a foodstuff, preferably a bakery product or a substance  
CC (e.g. a dough) for making the bakery product. Wild type xylanase or  
CC mutant xylanase is useful for preparing a dough that is less sticky  
CC than a dough comprising a fungal xylanase. The xylanase inhibitor is  
CC useful for screening high degree resistance xylanases for dough  
CC preparation. The xylanase is also useful for preparing a non-sticky  
CC dough. A combination of xylanase and the inhibitor is useful for  
CC calibrating and/or determining the quantity of inhibitor in a wheat  
CC flour sample.  
SQ Sequence 31 AA:  
Query Match 100.0%; Score 145; DB 21; Length 31;  
Best Local Similarity 100.0%; Pred. No. 2.1e-13;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LASLPRGSTGVAGLANSGLALPAQVASAOK 31  
DB 1 LASLPRGSTGVAGLANSGLALPAQVASAOK 31  
RESULT 2  
AAEL14690  
ID AAEL14690 standard; peptide: 31 AA.  
XX  
XX AAEL14690;  
AC  
XX  
XX 21-AUG-2002 (first entry)  
DT  
XX  
DE Wheat flour xylanase inhibitor B chain Lys-C digested fragment #1.  
XX  
XX Refrigerated dough; syrruping; arabinoxylan; bakery product; bread;  
KW pizza base; cake; biscuit; wheat; flour; xylanase inhibitor.  
XX  
XX Triticum aestivum.  
OS  
XX  
XX WO200152657-A1.  
PN  
XX  
XX 26-JUL-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-1B00168.  
PF  
XX  
XX 18-JAN-2000; 2000GB-0001136.  
PR  
XX  
XX (DANI-) DANISCO AS.  
PA  
XX  
XX Poulsen CH, Sorensen JF;  
PI  
XX  
XX WPI; 2001-457446/49.  
DR  
XX  
XX Production of refrigerated dough with reduced syrruping, useful in  
PT production of bakery products such as bread, comprises admixing cereal  
PT flour, water and protein that prevents enzymatic degradation of  
PT arabinoxylan in the cereal flour -  
PT  
XX  
XX Disclosure; Page 23; 26pp; English.  
PS  
XX The invention relates to a process for producing refrigerated dough  
CC with reduced 'syrruping' (precipitation of liquid on the dough surface  
CC because of a reduction in water holding capacity caused by the breakdown  
CC of arabinoxylan over time). The process comprises admixing cereal flour  
CC and water with a protein that reduces/prevents enzymatic degradation of  
CC arabinoxylan in the cereal flour. The preferred protein is a xylanase  
CC inhibitor. The method is useful to produce refrigerated dough in which  
CC syrruping is reduced or eliminated. Refrigerated dough is typically  
CC stored for long periods to enable fresh baked products (e.g. bread,  
CC pizza bases, cakes, biscuits) to be produced as required. It may also be

CC possible to produce dough or baked products tailored to specific  
CC requirements using the method by the use of specific proteins/protein  
CC combinations. The present sequence is wheat flour  
CC endo-beta-1,4-xylanase inhibitor B chain Lys-C digested fragment.  
XX  
SQ Sequence 31 AA:  
Query Match 100.0%; Score 145; DB 22; Length 31;  
Best Local Similarity 100.0%; Pred. No. 2.1e-13;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LASLPRGSTGVAGLANSGLALPAQVASAOK 31  
DB 1 LASLPRGSTGVAGLANSGLALPAQVASAOK 31  
RESULT 3  
AAU07395  
ID AAU07395 standard; protein: 31 AA.  
XX  
XX AAU07395;  
AC  
XX  
XX 18-DEC-2001 (first entry)  
DT  
XX  
DE Bacillus subtilis xylanase inhibitor #4.  
XX  
XX Bacillus subtilis xylanase inhibitor #4.  
DE  
XX  
XX Xylanase; plant cell wall; baking; cereal; starch production; wood;  
KW wood pulp bleaching; hemicellulose; arabinoxylan; foodstuff; inhibitor.  
XX  
XX Bacillus subtilis.  
OS  
XX  
XX WO200166711-A1.  
PN  
XX  
XX 13-SEP-2001.  
PD  
XX  
XX 08-MAR-2001; 2001WO-1B00426.  
PF  
XX  
XX 08-MAR-2000; 2000GB-0005585.  
PR  
XX  
XX 27-JUN-2000; 2000GB-0015751.  
PR  
XX  
XX (DANI-) DANISCO AS.  
PA  
XX  
XX Sibbesen O, Sorensen JF;  
PI  
XX  
XX WPI; 2001-596834/67.  
DR  
XX  
XX Novel variant xylanase polypeptide or its fragment useful for degrading  
PT or modifying plant cell wall, comprises amino acid modifications such  
PT that the polypeptide has altered sensitivity to xylanase inhibitor -  
PT  
XX  
XX Disclosure; Page 63; 70pp; English.  
PS  
XX The invention relates to a variant xylanase polypeptide (I) or its  
CC fragment having xylanase activity, comprising one or more amino acid  
CC modifications such that (I) or its fragment has an altered sensitivity to  
CC a xylanase inhibitor as compared with the parent xylanase enzyme. (I) or  
CC its coding sequence (II) is useful for degrading or modifying plant cell  
CC wall or for processing a plant material by contacting the plant cell wall  
CC or plant material with (I) or (II). (I) is useful for modifying plant  
CC materials, and in baking, processing cereals, starch production,  
CC processing wood and enhancing the bleaching of wood pulp. (I)  
CC is useful for altering the viscosity derived from the presence of  
CC hemicellulose or arabinoxylan in a solution or system comprising plant  
CC cell wall material. (I) is useful for preparing a foodstuff such as  
CC bread, pretzels, tortillas, cakes, cookies, biscuits or crackers.  
CC The present sequence represents the amino acid sequence of Bacillus  
CC subtilis xylanase inhibitor #4 as described in the method of the  
CC invention.  
SQ Sequence 31 AA:  
Query Match 100.0%; Score 145; DB 22; Length 31;  
Best Local Similarity 100.0%; Pred. No. 2.1e-13;



Matches 31: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LLASIPRGSTGVAGLSGLALPAQVASAOK 31  
 |||  
 Db 1 LLASIPRGSTGVAGLSGLALPAQVASAOK 31

RESULT 4  
 ID AAU75815 standard; Protein: 370 AA.  
 AC AAU75815;  
 DE 23-APR-2002 (first entry)

Wheat L endoxylanase inhibitor, TAXI I, partial sequence TAXI-I.0.  
 KW Wheat; TAXI-I: L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX Triticum aestivum cultivar Soissons.  
 PN WO200198474-A1.  
 PD 27-DEC-2001.  
 XX 21-JUN-2001; 2001WO-BE00106.  
 PF 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX (LEUV-) LEUVEN RES & DEV.  
 PA Delcour J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR N-PSDB: ABK13672.  
 DR WPI: 2002-114579/15.  
 XX Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX Claim 127; Page 57; 127pp; English.

The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitor  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable

CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial wheat TAXI-I (T. aestivum L endoxylanase  
 CC inhibitor).

Sequence 370 AA:  
 SQ Query Match 100.0%; Score 145; DB 23; Length 370;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-12;  
 Matches 31: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LLASIPRGSTGVAGLSGLALPAQVASAOK 31  
 |||  
 Db 118 LLASIPRGSTGVAGLSGLALPAQVASAOK 148

RESULT 5  
 ID AAU75816 standard; Protein: 370 AA.  
 AC AAU75816;  
 DE 23-APR-2002 (first entry)

Wheat L endoxylanase inhibitor, TAXI I, partial sequence TAXI-I.02.  
 KW Wheat; TAXI-I: L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX Triticum aestivum cultivar Estica.  
 OS WO200198474-A1.  
 PN 27-DEC-2001.  
 PD 21-JUN-2001; 2001WO-BE00106.  
 PF 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX (LEUV-) LEUVEN RES & DEV.  
 PA Delcour J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR N-PSDB: ABK13673.  
 DR WPI: 2002-114579/15.  
 XX Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX Claim 127; Page 57; 127pp; English.

The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the

CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malling of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial wheat TAXI-I (T. aestivum L endoxylanase  
 CC inhibitor).

XX Sequence 370 AA:

Query Match 100.0%; Score 145; DB 23; Length 370;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-12;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LLASLPRGSTGVAGLANSGLALPAQVASAOK 31  
 |||||  
 Db 118 LLASLPRGSTGVAGLANSGLALPAQVASAOK 148

RESULT 6  
 AAU75813  
 ID AAU75813 standard; Protein; 381 AA.

XX AAU75813;

DT 23-APR-2002 (first entry)

XX Wheat L endoxylanase inhibitor, TAXI I, variant #2.

XX Wheat, TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KM xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KM immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KM straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KM noodle; animal feed; starch separation; maize processing; malling;  
 KM plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 OS Triticum aestivum.

XX Friticum aestivum.

FT key Location/Qualifiers

FT Misc-difference 95 /label= Unknown

FT Misc-difference 98 /label= Unknown

FT Misc-difference 101 /label= Unknown

FT Misc-difference 110 /label= Unknown

FT Misc-difference 333 /label= Unknown

PN WO200198474-A1.

XX 27-DEC-2001.

PF 21-JUN-2001: 2001WO-BE00106.

XX 22-JUN-2000: 2000GB-0015296.

PR 25-JAN-2001: 2001GB-0002018.

PR 26-JAN-2001: 2001GB-0002194.

PR 16-MAR-2001: 2001GB-0006564.

PR 21-MAY-2001: 2001GB-0012328.

XX (LEUV-) LEUVEN RES & DEV.

PI Delcourt J., Debysse W., Gebuere K., Goesaert H., Pierens K., Robben J.

PI Van Campenhout S.

DR WPI: 2002-114579/15.

PS Claim 127; Page 51; 127pp; English.

CC The invention relates to separating and/or isolating inhibitors of

CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises

CC screening the inhibition activity by using two or more enzymes during the

CC separation and/or isolation steps that allow to distinguish inhibitors of

CC different specificity or by using an affinity chromatographic step with

CC immobilised enzymes and/or antibodies against inhibitors. Also

CC included are an isolated nucleic acid molecule encoding an inhibitor

CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,

CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,

CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the

CC nucleic acid, an expression system transformed with the nucleic

CC acid, a host organism transformed with the nucleic acid, the inhibitory

CC proteins encoded by the nucleic acids and modulators of the proteins.

CC A recombinant protein, glycoprotein or polypeptide or microorganisms,

CC plant or plant materials transformed with the nucleic acid are useful

CC for the formation of an endoxylanase-inhibitor complex, screening

CC endoxylanases that are totally, less or not inhibited by the inhibitors,

CC reducing syruping in refrigerated dough compositions, affecting the

CC relative affinity and/or relative hydrolysis specificity and/or relative

CC hydrolysis rate versus water-extractable and/or water-unextractable

CC arabinoxylans of endoxylanases such as by the formation of an

CC endoxylanase/inhibitor complex, improving the malling of cereals such as

CC barley, sorghum and wheat and/or the production of beer, improving the

CC production and/or quality of baked or extruded cereal products such as

CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a variant of wheat TAXI-I (T. aestivum L endoxylanase  
 CC inhibitor).  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.

XX Sequence 381 AA:

Query Match 100.0%; Score 145; DB 23; Length 381;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-12;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LLASLPRGSTGVAGLANSGLALPAQVASAOK 31  
 |||||

Db 129 LLASLPRGSTGVAGLANSGLALPAQVASAOK 159

RESULT 7

AAU75812

ID	AAU75812 standard; Protein; 381 AA.
XX	
AC	AAU75812;
XX	
DT	23-APR-2002 (first entry)
XX	
DE	Wheat L endoxylanase inhibitor, TAXI I, variant #1.
XX	
KW	Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;
KW	xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;
KW	immobilised enzyme; enzyme; dough syruiping; cereal product; beer;
KW	straight dough; sponge dough; Chorleywood bread; biscuits; pasta;
KW	noodle; animal feed; starch separation; maize processing; malting;
KW	plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.
XX	
OS	Triticum aestivum.
XX	
FT	Key Location/Qualifiers
FT	Misc-difference 95 /label= Unknown
FT	Misc-difference 98 /label= Unknown
FT	Misc-difference 101 /label= Unknown
FT	Misc-difference 110 /label= Unknown
FT	Misc-difference 333 /label= Unknown
FT	Misc-difference /label= Unknown
XX	
PN	WO200198474-A1.
XX	
PD	27-DEC-2001.
XX	
PF	21-JUN-2001; 2001WO-BE00106.
XX	
PR	22-JUN-2000; 2000GB-0015296.
PR	25-JAN-2001; 2001GB-0002018.
PR	26-JAN-2001; 2001GB-0002194.
PR	16-MAR-2001; 2001GB-0006564.
PR	21-MAY-2001; 2001GB-0012328.
XX	
PA	(LEUV-) LEEVEN RES & DEV.
XX	
PI	DeJoucou J, Debysse W, Gebuere K, Goesaert H, Fierens K, Robben J;
PI	Van Campenhout S;
XX	
DR	WPI: 2002-114579/15.
XX	
PT	Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or
PT	beta-glucanolytic enzymes comprises using endoxylanases during
PT	screening for inhibition activity or affinity chromatography with
PT	immobilised enzymes -
XX	
PS	Claim 127; Page 50-51; 127pp; English.
XX	
CC	The invention relates to separating and/or isolating inhibitors of
CC	cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising
CC	screening the inhibition activity by using two or more enzymes during
CC	separation and/or isolation steps that allow to distinguish inhibitors of
CC	different specificity or by using an affinity chromatographic step with
CC	immobilised enzymes and/or antibodies against inhibitors. Also
CC	included are an isolated nucleic acid molecule encoding an inhibitor
CC	which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,
CC	alpha-L-arabino-furanosidase and/or other cellulose, xylan,
CC	arabinoxylan or beta-glucan degrading enzymes, a vector comprising the
CC	nucleic acid, an expression system transformed with the nucleic
CC	acid, a host organism transformed with the nucleic acid, the inhibitory
CC	proteins encoded by the nucleic acids and modulators of the proteins.
CC	A recombinant protein, glycoprotein or polypeptide or microorganisms,
CC	plant or plant materials transformed with the nucleic acid are useful
CC	for the formation of an endoxylanase-inhibitor complex, screening
CC	endoxylanases that are totally, less or not inhibited by the inhibitors,
CC	reducing syruiping in refrigerated dough compositions, affecting the

CC	relative affinity and/or relative hydrolysis specificity and/or relative
CC	hydrolysis rate versus water-extractable and/or water-unextractable
CC	arabinoxyans of endoxylanases such as by the formation of an
CC	endoxylanase/inhibitor complex, improving the malting of cereals such as
CC	barley, sorghum and wheat and/or the production of beer, improving the
CC	production and/or quality of baked or extruded cereal products such as
CC	straight dough, sponge dough, Chorleywood bread, breakfast cereals,
CC	biscuits, pasta and noodles, animal feed stuff, improving the production
CC	of starch derived syrups, sorbitol, xylitol, maltose and/or xyloitol, wheat gluten
CC	starch separation and production, improving maize processing,
CC	plant disease resistance and nutraceutical and/or pharmaceutical
CC	applications, improving paper and pulp technologies. The present
CC	sequence is a variant of wheat TAXI-I (' <i>aestivum</i> L endoxylanase
CC	inhibitor').
CC	Note: Variant amino acids are highlighted in the specification but
CC	no wild-type sequence is shown for comparison.
XX	
SQ	Sequence      381 AA;
	Query Match                  96.6%; Score 140; DB 23; Length 381;
	Best Local Similarity        96.8%; Pred.No.1.7e-11;
	Matches     30; Conservative    1; Mismatches    0; Indels        0; Gaps        0;
OY	1 LLASIPRGSTGVAGLANSGLALPAOVASAK 31
D6	129 LLASIPRGSTGVAGLADSLPALPAOVASAK 159
RESULT 8	
AU075814	
ID	AAU75814 standard; Protein: 381 AA.
XX	
AC	AAU75814;
XX	
DT	23-APR-2002 (first entry)
XX	
DE	Wheat L endoxylanase inhibitor, TAXI I, microheterogenic variant.
XX	
KM	Wheat: TAXI-I, L endoxylanase inhibitor; cellulolytic enzyme inhibitor;
KM	xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;
KM	immobilised enzyme; enzyme; dough syruping; cereal product; beet; plant;
KM	straight dough; sponge dough; Chorleywood bread; biscuits; pasta;
KM	noodle; animal feed; starch separation; maize processing; malting;
KM	plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.
XX	
OS	Triticum aestivum.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 95
FT	/label= Unknown
FT	Misc-difference 98
FT	/label= Unknown
FT	Misc-difference 101
FT	/label= Unknown
FT	Misc-difference 110
FT	/label= Unknown
FT	Misc-difference 145
FT	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 183
FT	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 232
FT	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 275
FT	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 282
FT	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 333
FT	/label= Unknown
XX	
PN	WO200198474-A1.
XX	
PD	27-DEC-2001.
XX	

PF 21-JUN-2001; 2001WO-BE00106.  
 XX  
 PR 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.  
 XX  
 PI Delcours J, Debysier W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR WPI; 2002-114579/15.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes -  
 XX  
 PS Claim 127; Page 51; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruup in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a microheterogenic variant of wheat TAXI-I (T. aestivum L  
 CC endoxylanase inhibitor)  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.  
 XX  
 SO Sequence 381 AA:  
 Query Match 95.2%; Score 138; DB 23; Length 381;  
 Best Local Similarity 96.8%; Pred. No. 3.2e-11;  
 Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LLASLPRGSTGVAGLANSGLALPAOVASAK 31  
 DB 129 LLASLPRGSTGVAGLANSGLALPAOVASAK 159  
 RESULT 9  
 AAU75818  
 ID AAU75818 standard; Protein: 277 AA.  
 XX  
 AC AAU75818;

XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE L endoxylanase inhibitor, ATXI-II, partial sequence ATXI-II.01.  
 XX  
 KW ATXI-II: L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor; ds;  
 KW immobilised enzyme; enzyme; dough syruuping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Aegilops tauschii variety Acc2220051.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 4 /note= "Encoded by CTC"  
 FT  
 XX  
 PN WO200198474-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 21-JUN-2001; 2001WO-BE00106.  
 XX  
 PR 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.  
 XX  
 PI Delcours J, Debysier W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR WPI; 2002-114579/15.  
 DR N-PSDB: ABK13675.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes -  
 XX  
 PS Claim 127; Page 58; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruuping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC

CC		plant disease resistance and nutraceutical and/or pharmaceutical applications, improving paper and pulp technologies. The present sequence represents a partial Aegilops tauschii ATXI-II (Aegilops tauschii L endoxylanase inhibitor).
CC		
CC		
CC		
XX		
SO	Sequence	277 AA;
OY		
DB		
Db	118	LILASLRGSGVAGLGGSLAPPAQVAASKK 148                              1 LILASLRGSGVAGLGGSLAPPAQVAASKK 31
RESULT 10		
ID	AAU75819	standard; Protein: 287 AA.
AC	AAU75819:	
XX		
DT	23-APR-2002	(first entry)
DE		
XX		
XX		
KW	Wheat; TAXI-III; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;	
KM	xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;	
KW	immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;	
KM	straight dough; sponge dough; Chorleywood Bread; biscuits; pasta;	
KW	noodle; animal feed; starch separation; maize processing; malting;	
RN	plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.	
XX		
OS	Triticum aestivum cultivar Solissons.	
XX		
PN	WO200198474-A1.	
PD		
XX		
PD	27-DEC-2001.	
XX		
PF	21-JUN-2001; 2001WO-BE00106.	
XX		
PR	22-JUN-2000; 2000GB-0015296.	
PR	25-JAN-2001; 2001GB-0002018.	
PR	26-JAN-2001; 2001GB-0002194.	
PR	16-MAR-2001; 2001GB-0006564.	
PR	21-MAY-2001; 2001GB-0012328.	
XX		
PA	(LEUV-) LEUVEN RES & DEV.	
XX		
PI	Delaourt J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J,	
PI	Van Campenhout S;	
DR	N-PSSB; ABK13676.	
XX		
MPT	WPt: 2002-114579/15.	
PT		
PT		
PT		
XX		
PS	Claim 127; Page 58; 127pp; English.	
XX		
XX	The invention relates to separating and/or isolating inhibitors of	
CC	cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises	
CC	screening the inhibition activity by using two or more enzymes during the	
CC	separation and/or isolation steps that allow to distinguish inhibitors of	
CC	different specificity or by using an affinity chromatography step with	
CC	immobilised enzymes and/or antibodies against inhibitors. Also	
CC	included are an isolated nucleic acid molecule encoding an inhibitor	
CC	which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,	
CC	alpha-L-arabinofuranosidase and/or other cellulose, xylan,	
CC	arabinoxylan or beta-glucan degrading enzymes, a vector comprising the	
CC	nucleic acid, an expression system transformed with the nucleic	
CC		

	acid, a host organism transformed with the nucleic acid, the inhibitory
CC	proteins encoded by the nucleic acids and modulators of the proteins.
CC	A recombinant protein, glycoprotein or polypeptide or microorganisms,
CC	plant or plant materials transformed with the nucleic acid are useful
CC	for the formation of an endoxylanase-inhibitor complex, screening
CC	endoxyanases that are totally, less or not inhibited by the inhibitors,
CC	reducing syruiping in refrigerated dough compositions, affecting the
CC	relative affinity and/or relative hydrolysis specificity and/or relative
CC	hydrolysais rate versus water-extractable and/or water-unextractable
CC	arabinoxyans of endoxylanases such as by the formation of an
CC	endoxylanase/inhibitor complex, improving the malting of cereals such as
CC	bareley, sorgnum and wheat and/or the production of beer, improving the
CC	production and/or quality of baked or extruded cereal products such as
CC	straight dough, sponge dough, Chorlewood bread, breakfast cereals,
CC	discults, pasta and noodles, animal feed stuff, improving the production
CC	of starch derived syrups, sorbitol, xylose and/or xyllitol, wheat gluten
CC	starch separation and production, improving maize processing,
CC	plant disease resistance and nutraceutical and/or pharmaceutical
CC	applications, improving paper and pulp technologists. The present
CC	sequence represents a partial wheat TAxI-III ( <i>T. aestivum L</i> ) endoxylanase
CC	inhibitor).
XX	
SQ	Sequence      287 AA;
	Query Match                  76.6%; Score 111; DB 23; Length 287;
	Best Local Similarity        77.4%; Pred. No. 1.5e-07;
Matches	24; Conservative     3; Mismatches     4; Indels        0; Gaps        0;
Oy	1 LLASLPRGSTGVAGLANSGLAPAOVYASOK 31    :::  ::             ::
Dd	120 LLESIPGGSVGAVGLAGSDLAIPAQVASERK 150
<hr/>	
RESULT 11	
ID	AAU75824 standard; Protein: 195 AA.
AC	AAU75824;
DT	23-APR-2002 (first entry)
DE	Barley L endoxylanase inhibitor, HVXI I, variant #1.
KX	
KM	Barley; HVXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;
KW	xylanolytic enzyme inhibitor; beta-glucanoalytic enzyme inhibitor;
KW	immobilized enzyme; enzyme; dough syruping; cereal product; beer; plant;
KW	straight dough; sponge dough; Chorlewood bread; Discults; pasta;
KW	noodle; animal feed; starch separation; maize processing; maling;
XO	plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.
OS	Hordeum vulgare.
FH	Key Location/Qualifiers
FT	Misc-difference 72 /Label= Unknown
FT	Misc-difference 185 /Label= Unknown
PN	WO200198474-A1.
PD	
PE	27.-DEC-2001.
PR	21-JUN-2001; 2001WO-BE00106.
PR	
PR	22-JUN-2000; 2000GB-0015296.
PR	25-JAN-2001; 2001GB-0002018.
PR	26-JAN-2001; 2001GB-0002194.
PR	16-MAR-2001; 2001GB-0006564.
PR	21-MAY-2001; 2001GB-0012328.
PA	(LEUV-) LEUVEN RES & DEV.
PI	Delcourt J., Debysier W., Gebruens K., Goesnaert H., Pierssens R., Robben J.;

PI Van Camphenout S;  
 XX  
 DR WPI; 2002-114579/15.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprising using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 63; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC arabinoxylan rate versus water-extractable and/or water-unextractable  
 CC endoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a variant of barley HVXI-I (H. vulgare L endoxylanase  
 CC inhibitor).  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.  
 CC  
 XX  
 SO Sequence 185 AA;  
 Query Match 53.8%; Score 78; DB 23; Length 185;  
 Best Local Similarity 89.5%; Pred. No. 0.0038;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 13 AGLANSGALPQVVASAK 31  
 DB 1 AGFAGSGALPQVVASAK 19  
 ID AAU75825 standard; Protein: 185 AA.  
 AC AAU75825;  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Barley L endoxylanase inhibitor, HVXI I, variant #2.  
 XX  
 XX Barley; HVXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

XX OS Hordeum vulgare.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 72  
 FT Misc-difference 185  
 FT Misc-difference 185 /label= unknown  
 XX  
 XX W0200198474-A1.  
 PN  
 XX  
 PD 27-DEC-2001.  
 XX  
 XX 21-JUN-2001; 2001WO-BE00106.  
 XX  
 PF 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 XX (LEUV-) LEUVEN RES & DEV.  
 PA  
 PI Delcours J, Debysse W, Gebuere K, Goesaert H, Flierens K, Robben J;  
 PI Van Camphenout S;  
 XX  
 XX WPI; 2002-114579/15.  
 DR  
 XX  
 XX Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprising using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 63; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a variant of barley HVXI-I (H. vulgare L endoxylanase  
 CC inhibitor).  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.  
 CC  
 XX  
 SO Sequence 185 AA;  
 Query Match 53.8%; Score 78; DB 23; Length 185;

Best Local Similarity 89.5%; Pred. No. 0.0038;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 13 AGLANGSLALPAQVSAQK 31  
||| |||||  
Db 1 AGFAGSGLALPAQVSAQK 19

## RESULT 13

AAU75826  
ID AAU75826 standard; Protein: 185 AA.

AC AAU75826;

DT 23-APR-2002 (first entry)

DE Barley L endoxylanase inhibitor, HVXI I, microheterogenic variant.

KW Barley; HVXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;

KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;

KW immobilised enzyme; enzyme; dough syruing; cereal product; beer; plant;

KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;

KW noodle; animal feed; starch separation; maize processing; malting;

OS plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

OS Synthetic.

FT Key Location/Qualifiers

FT Misc-difference 72 /Label= Unknown

FT Misc-difference 128 /Label= Pro, Ala, Asp, Leu, Gly, Asn, Cys

FT Misc-difference 132 /Label= Pro, Ala, Asp, Leu, Gly, Asn, Cys

FT Misc-difference 134 /Label= Pro, Ala, Asp, Leu, Gly, Asn, Cys

FT Misc-difference 148 /Label= Pro, Ala, Asp, Leu, Gly, Asn, Cys

FT Misc-difference 185 /Label= Unknown

XX WO200198474-A1.

XX 27-DEC-2001.

XX 21-JUN-2001; 2001WO-BE00106.

XX 22-JUN-2000; 2000GB-0015296.

XX 25-JAN-2001; 2001GB-0002018.

XX 26-JAN-2001; 2001GB-0002194.

XX 16-MAR-2001; 2001GB-0006564.

XX 21-MAY-2001; 2001GB-0012328.

XX (LEUV-) LEUVEN RES & DEV.

XX Delcourt J, Debysier W, Gebruers K, Goesaert H, Flierens K, Robben J;

XX Van Campenhout S;

XX WPI: 2002-114579/15.

PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or

CC beta-glucanolytic enzymes comprises using endoxylanases during

CC screening for inhibition activity or affinity chromatography with

CC immobilised enzymes

CC Claim 127; Page 63; 127pp; English.

CC The invention relates to separating and/or isolating inhibitors of

CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises

CC screening the inhibition activity by using two or more enzymes during the

CC separation and/or isolation steps that allow to distinguish inhibitors of

CC different specificity or by using an affinity chromatographic step with

CC immobilised enzymes and/or antibodies against inhibitors. Also

CC included are an isolated nucleic acid molecule encoding an inhibitor

CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,

CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,

CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the

CC nucleic acid, a host organism transformed with the nucleic acid, the inhibitory

CC proteins encoded by the nucleic acids and modulators of the proteins.

CC A recombinant protein, glycoprotein or polypeptide or microorganisms,

CC plant or plant materials transformed with the nucleic acid are useful

CC for the formation of an endoxylanase-inhibitor complex, screening

CC endoxylanases that are totally, less or not inhibited by the inhibitors,

CC reducing syruing in refrigerated dough compositions, affecting the

CC relative affinity and/or relative hydrolysis specificity and/or relative

CC hydrolysis rate versus water-extractable and/or water-unextractable

CC arabinoxylans of endoxylanases such as by the formation of an

CC endoxylanase/inhibitor complex, improving the malting of cereals such as

CC barley, sorghum and wheat and/or the production of beer, improving the

CC production and/or quality of baked or extruded cereal products such as

CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,

CC biscuits, pasta and noodles, animal feed stuff, improving the production

CC of starch derived syrups, sorbitol, xylitol, wheat gluten

CC starch separation and production, improving maize processing,

CC plant disease resistance and nutraceutical and/or pharmaceutical

CC applications, improving paper and pulp technologies. The present

CC sequence is a microheterogenic variant of barley HVXI-I (H. vulgare L

CC endoxylanase inhibitor).

CC Note: Variant amino acids are highlighted in the specification but

CC no wild-type sequence is shown for comparison.

XX SQ Sequence 185 AA;

Query Match 53.8%; Score 78; DB 23; Length 185;

Best Local Similarity 89.5%; Pred. No. 0.0038;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 13 AGLANGSLALPAQVSAQK 31  
||| |||||  
Db 1 AGFAGSGLALPAQVSAQK 19

## RESULT 14

AAG42491  
ID AAG42491 standard; Protein: 433 AA.

XX AAG42491;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 52999.

XX Protein identification; signal transduction pathway; metabolic pathway;

XX hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

PR 19-APR-1999;	99US-0130077.
PR 21-APR-1999;	99US-0130449.
PR 23-APR-1999;	99US-0130510.
PR 23-APR-1999;	99US-0130891.
PR 28-APR-1999;	99US-0131449.
PR 30-APR-1999;	99US-0132048.
PR 30-APR-1999;	99US-0132407.
PR 04-MAY-1999;	99US-0132484.
PR 05-MAY-1999;	99US-0132485.
PR 06-MAY-1999;	99US-0132487.
PR 07-MAY-1999;	99US-0132863.
PR 11-MAY-1999;	99US-0134256.
PR 14-MAY-1999;	99US-0134218.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134321.
PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135124.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 18-JUN-1999;	99US-0139763.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139899.
PR 23-JUN-1999;	99US-0140353.
PR 23-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 05-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157863.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.





PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153738.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154479.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 44.1%; Score 64; DB 21; Length 433;  
Best Local Similarity 41.4%; Pred. No. 0.89; Mismatches 10; Indels 0; Gaps 0;  
Matches 12; Conservative 7;

Oy 1 LLASLPKSGTGVAGLANSGLAPQVASA 29  
Db 165 LKGLAKGTGVMAGMGRNIGLPSQFAA 193

Search completed: May 20, 2003, 16:14:41  
Job time : 34.7253 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:11:22 ; Search time 12.6044 Seconds  
(without alignments)  
236.439 Million cell updates/sec

Title: US-09-869-155-16

Perfect score: 145  
Sequence: 1 LLASLPGRSGTVAGLANSGLALPAQVASAOK 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR\_73:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	46.9	433	2	T14329
2	66	45.5	427	1	S06750
3	64	44.1	433	2	F86163
4	62	42.8	434	2	G86163
5	55	37.9	456	1	TVHDMC
6	55	37.9	460	2	S11558
7	53	37.9	464	1	TVHDM2
8	53	36.6	330	2	T29640
9	53	36.6	520	2	H64510
10	52	35.9	523	2	T40998
11	51	35.2	406	2	A53010
12	50.5	34.8	503	2	G75262
13	50	34.5	468	2	T34679
14	50	34.5	1114	2	T18197
15	49.5	34.1	310	2	JN0490
16	49.5	34.1	672	2	H83315
17	49.5	34.1	1443	2	G75393
18	49.5	34.1	1752	2	A45407
19	49	33.8	135	2	D64339
20	49	33.8	377	2	C91270
21	49	33.8	377	2	C91270
22	49	33.8	652	1	S39409
23	48.5	33.4	157	2	AC2775
24	48.5	33.4	162	2	E97555
25	48.5	33.4	485	2	G65051
26	48.5	33.4	485	2	B85920
27	48.5	33.4	485	2	C91075
28	48.5	33.4	1222	2	G59100
29	48	33.1	189	2	AC3420

30	48	33.1	244	2	F70834	hypothetical prote
31	48	33.1	473	2	C87325	nitrate transpor
32	48	33.1	601	1	B56564	transcription fac
33	48	33.1	608	2	T03476	conserved hypoth
34	48	33.1	613	2	T00077	gag-like protein
35	48	33.1	719	2	T00266	hypothetical prote
36	48	33.1	750	2	H70711	probable multib prot
37	48	33.1	839	2	F75518	hypothetical prote
38	48	33.1	1046	2	F71432	hypothetical prote
39	47.5	32.8	238	1	LKRIB2	clathrin light cha
40	47.5	32.8	446	2	T50786	nucleosid DNA-bind
41	47.5	32.8	3788	2	T30851	lysosomal traffick
42	47	32.4	72	2	A11026	hypothetical prote
43	47	32.4	216	2	I51234	paired box transcr
44	47	32.4	324	2	H83127	probable dehydroge
45	47	32.4	337	2	F83877	acetoin dehydrogen

## ALIGNMENTS

RESULT 1  
T14329  
dermal glycoprotein precursor, extracellular - carrot (fragment)  
C:Species: Daucus carota (carrot)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Nov-2000  
C:Accession: T14329  
R:Satoh, S.; Sturm, A.; Fujii, T.; Chrispeels, M.J.  
A:Title: CDNA cloning of an extracellular dermal glycoprotein of carrot and its expression  
A:Reference number: Z17987  
A:Accession: T14329  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-433 <SAT>  
A:Cross-references: EMBL:D14550; NID:d1004922; PID:d1003923  
A:Experimental source: subspecies US-Harumakigosun; root  
C:Genetics:  
A:Gene: EDGP1  
C:Superfamily: conglutinin gamma  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-433/Product: dermal glycoprotein, extracellular #status predicted <MAT>

Query Match 46.9% Score 68; DB 2; Length 433;  
Best Local Similarity 51.7%; Pred. No. 0.21;  
Matches 15; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 1 LLASLPGRSGTVAGLANSGLALPAQVASA 29  
DB 163 LLONLASGVVGMAGLGRTRIALPSQFASA 191

RESULT 2  
S06750  
basic 7S globulin precursor - soybean  
N:Alternate names: insulin-binding protein  
C:Species: Glycine max (soybean)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S06750; S00218; A61003; JN0311; JN0312  
R:Kagawa, H.; Hirano, H.  
A:Title: Sequence of a cDNA encoding soybean basic 7S globulin.  
A:Reference number: S06750; NUID:90067863; PMID:2587227  
A:Accession: S06750  
A:Molecule type: mRNA  
A:Residues: 1-427 <KAI>  
A:Cross-references: EMBL:X16469; NID:g18542; PID:CA34489.1; PID:g18543  
R:Kagawa, H.; Yamauchi, F.; Hirano, H.  
FEBS Lett. 226, 145-149, 1987  
A:Title: Soybean basic 7S globulin represents a protein widely distributed in legume  
A:Reference number: S00218  
A:Accession: S00218  
A:Molecule type: protein

A:Residues: 276-293,'X',295-304,'FX',423-427,'X','TVPGPGCLCPNONGVTSLGPMMXMPAR','QL',356,42,'AX' <KAG>  
 R:Hirano, H.; Watanabe, T.  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
 C:Accession: G86163  
 A:Title: Microsequencing of proteins electrotransferred onto immobilizing matrices from  
 A:Reference number: A61003; MUID:91031409; PMID:2226413  
 A:Accession: A61003  
 A:Molecule type: protein  
 A:Residues: 25-57,135-150,276-304,'X',306-308,'X',310-311,383-417 <HIR>  
 R:Barbashov, S.F.; Egorov, T.A.; Kochkina, V.M.  
 Bioorg. Khim. 17, 421-423, 1991  
 A:Title: Isolation and characterization of soybean insulin-binding protein.  
 A:Reference number: JN0319; MUID:91291181; PMID:2064630  
 A:Accession: JN0311  
 A:Molecule type: protein  
 A:Residues: 'VPIPDHKTN',27-33,'X',35-42,'XX',45,'X',47,'X',49,'AX',52 <BAR>  
 A:Note: 20-Thr,21-Asn, 25-Pro and 50-Asn were also found  
 A:Accession: JN0312  
 A:Molecule type: protein  
 A:Residues: 'S',99,'IV',102,'SX',105,'C',286-299,'XX',302,'XX',305 <BA2>  
 A:Note: 102-Gly and 296-Lys were also found  
 C:Comment: This protein specifically binds to bovine insulin.  
 C:Superfamily: conglutin gamma  
 C:Keywords: heterodimer; seed; storage protein  
 F:1-4/Domain: signal sequence #status predicted <SIG>  
 F:25-275/Product: basic 7S globulin light chain #status experimental <MAT1>  
 F:275-427/Product: basic 7S globulin light chain #status experimental <MAT2>

Query Match 45.5% Score 66; DB 1: Length 427;  
 Best Local Similarity 54.2%; Pred. No. 0.37;  
 Matches 13; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 5 LPRGSGVAGLANSGALPAQVSA 28  
 DB 175 LPRMGVAGLGHAPISLPNQLAS 198

RESULT 3  
 F66163  
 hypotetical protein F15K9.17 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001  
 C:Accession: F66163  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 Jensen, N.F.; Hughes, B.; Hulzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A66141; MUID:21016719; PMID:11130712  
 A:Molecule type: DNA  
 A:Accession: F66163  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-433 <STO>  
 A:Cross-references: GB:AE005172; NID:93850579; PIDN:AC72119.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: conglutin gamma

Query Match 44.1% Score 64; DB 2: Length 433;  
 Best Local Similarity 41.4%; Pred. No. 0.7;  
 Matches 12; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

OY 1 LLASLPRGSGVAGLANSGALPAQVSA 29  
 DB 165 LKGLAKGTVMGMGRHNIIGLPSQFAA 193

RESULT 4

G86163  
 hypotetical protein F15K9.16 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
 C:Accession: G86163  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
 Jensen, N.F.; Hughes, B.; Hulzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A66141; MUID:21016719; PMID:11130712  
 A:Accession: G86163  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-434 <STO>  
 A:Cross-references: GB:AE005172; NID:93850580; PIDN:AC72120.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: conglutin gamma

Query Match 42.8% Score 62; DB 2: Length 434;  
 Best Local Similarity 41.4%; Pred. No. 1.3;  
 Matches 12; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

OY 1 LLASLPRGSGVAGLANSGALPAQVSA 29  
 DB 166 LKGLAKGTVMGMGRHNIIGLPSQFAA 194

RESULT 5  
 TVHUMC  
 transforming protein N-myc (version 1) - human  
 C:Species: Homo sapiens (man)  
 C>Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 15-Oct-1999  
 C:Accession: A01355; I59514; S02249  
 R:Slamon, L.W.; Schwab, M.; Bishop, J.M.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 1772-1776, 1986  
 A:Title: Nucleotide sequence of the human N-myc gene.  
 A:Reference number: A01355; MUID:86149381; PMID:2869488  
 A:Accession: A01355  
 A:Molecule type: DNA  
 A:Residues: 1-456 <STA>  
 A:Cross-references: GB:M13241; NID:9189247; PIDN:AAA36371.1; PID:9386983  
 R:Slamon, D.J.; Boone, T.C.; Seeger, R.C.; Keith, D.F.; Chazin, V.; Lee, H.C.; Souza,  
 Science 232, 768-772, 1986  
 A:Title: Identification and characterization of the protein encoded by the human N-myc  
 A:Reference number: I59514; MUID:86179901; PMID:3008339  
 A:Accession: I59514  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 63-456 <RES>  
 A:Cross-references: GB:M13228; NID:9189243; PIDN:AAA36370.1; PID:9386982  
 R:Ibson, J.M.; Rabbits, P.H.  
 Oncogene 2, 399-402, 1988  
 A:Title: Sequence of a germ-line N-myc gene and amplification as a mechanism of activ  
 A:Reference number: S00945; MUID:88202932; PMID:2834684  
 A:Accession: S02249  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-354,'V',356-456 <IBS>  
 A:Cross-references: EMBL:Y00664; NID:935074; PIDN:CAA68678.1; PID:935076  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1988  
 C:Genetics:  
 A:Gene: GDB:MYCN; NMNC  
 A:Cross-references: GDB:120527; OMIM:164840  
 A:Map position: 2p24.3-2p24.3  
 A:Introns: 256/1  
 C:Superfamily: myc transforming protein; myc transforming protein homology



hypothetical protein MJEC108 - Methanococcus jannaschii plasmid pMB800  
 C:Species: Methanococcus jannaschii  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 29-Sep-1999  
 C:Accession: H64510  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.T.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Isori, J.D.; Sadow, P.W.; Hann, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C. A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii A:Reference number: A64300, MIMD:96337999; PMID:8688087  
 A:Accession: H64510  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-520 <BUIL>  
 A:Cross-references: GB:L77118; NID:g1500644; TIGR:MJEC108; PIDN:AAC37081.1; PID:g1522648  
 C:Genetics:  
 A:Map position: ECLREV7391-5829  
 A:Genome: Plasmid  
 A:Start codon: GTG  
 A:Note: this stable 58-kilobase pair plasmid is also designated ECL (large extrachromosomal C:Superfamily: conserved hypothetical protein MJ1429

Query Match 36.6%; Score 53; DB 2; Length 520;  
 Best Local Similarity 53.8%; Pred. No. 24;  
 Matches 14; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 1 LLASLRPGSTGVAGLANGLALPAQV 26  
 DB 475 LLQDLPLQSTGEATVGVSSLRPLALV 500

RESULT 10  
 T40998  
 hypothetical protein SPC1450.16c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T40998  
 R:Lvine, M.; Rajandream, M.A.; Barrell, B.G.; Voikaert, G. submitted to the EMBL Data Library, March 1999  
 A:Reference number: Z21962  
 A:Accession: T40998  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-513 <LYN>  
 A:Cross-references: EMBL:AL049559; PIDN:CAB40183.1; GSPDB:GN00068; SPDB:SPCC1450.16c  
 A:Experimental source: strain 972h.; cosmid c1450  
 C:Genetics:  
 A:Gene: SPDB:SPCC1450.16c  
 A:Map position: 3

Query Match 35.9%; Score 52; DB 2; Length 513;  
 Best Local Similarity 45.5%; Pred. No. 32;  
 Matches 15; Conservative 2; Mismatches 6; Indels 10; Gaps 1;

OY 1 LLASLRPGSTGV-----AGLANGLALP 23  
 DB 339 LLAKLPDGTSEVCTPKNTWPAAGLPMTGRSNP 371

RESULT 11  
 A53010  
 copper transport protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: Protein P9642.3; protein YPR124W  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Mar-2001  
 C:Accession: A53010; S69016; S48876  
 R:Dancis, A.; Yuan, D.S.; Hallie, D.; Askwith, C.; Elide, D.; Mochle, C.; Kaplan, J.; Klatz Cell 76, 393-402, 1994  
 A:Title: Molecular characterization of a copper transport protein in Saccharomyces cerevisiae  
 A:Reference number: A53010; MIMD:94123345; PMID:8293472  
 A:Accession: A53010  
 A:Molecule type: DNA

A:Residues: 1-406 <DAN>  
 A:Cross-references: GB:U02511; NID:g407947; PIDN:AA17369.1; PID:g407948  
 R:Johnson, D. submitted to the EMBL Data Library, November 1995  
 A:Description: The sequence of S. cerevisiae cosmid 9642.  
 A:Reference number: S69014  
 A:Accession: S69016  
 A:Molecule type: DNA  
 A:Residues: 1-406 <DAN>  
 A:Cross-references: EMBL:U040828; NID:g1066467; PIDN:AAB68064.1; PID:g1066470; MIPS:YP C:Genetics:  
 A:Gene: SGD:CYP1  
 A:Cross-references: SGD:S0006328; MIPS:YPR124W  
 A:Map position: 16R  
 A:Function:  
 A:Keywords: copper transport; transmembrane protein  
 F:156-172/Domain: transmembrane #status predicted <TM1>  
 F:255-271/Domain: transmembrane #status predicted <TM2>

Query Match 35.2%; Score 51; DB 2; Length 406;  
 Best Local Similarity 37.9%; Pred. No. 34;  
 Matches 11; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

OY 2 LASLRPGSTGVAGLANGLALPAQVASAO 30  
 DB 67 MASMSGSSMSGMSMSSTPTSSASAO 95

RESULT 12  
 G75262  
 hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C:Accession: G75262  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J. M.; Shen, M.; Vamathevan, J.D.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1. A:Reference number: A75250; MIMD:20036896; PMID:10567266  
 A:Accession: G75262  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-503 <HRT>  
 A:Cross-references: GB:AE002082; GB:AE000513; NID:g6460347; PIDN:AA12072.1; PID:g646 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR2528  
 A:Map position: 1

Query Match 34.8%; Score 50.5; DB 2; Length 503;  
 Best Local Similarity 42.9%; Pred. No. 50;  
 Matches 12; Conservative 5; Mismatches 8; Indels 3; Gaps 1;

OY 5 LPRGSTGVAGLANGLAL--PAOVASA 29  
 DB 81 IPAGMGITAGLSTGAAOOTSAPVPTAA 108

RESULT 13  
 T34679  
 probable protoporphyrinogen oxidase - Streptomyces coelicolor (fragment)

C:Species: Streptomyces coelicolor  
 C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 03-Nov-2000  
 C:Accession: T34679  
 R:Haris, D.; Taylor, K.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, May 1998  
 A:Reference number: Z21553  
 A:Accession: T34679  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-468 <HAR>



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:05:42 ; Search time 8.34615 Seconds  
(without alignments)

154.055 Million cell updates/sec

Title: US-09-869-155-16

Sequence: 1 LLASLPFGSTGVAGLANSLPALPAOVASAK 31

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	66	45.5	427 1	7SBG_SOYBN
2	55	37.9	460 1	MYCN_MARMO
3	55	37.9	464 1	MYCN_HUMAN
4	53	36.6	312 1	DIF1_CAEEL
5	53	36.6	520 1	VZ08_METJA
6	51	35.2	406 1	CTRL_YEAST
7	49.5	34.1	742 1	TBX3_HUMAN
8	49	33.8	124 1	SPEH_METJA
9	49	33.8	652 1	NOS2_PARDE
10	48.5	33.4	485 1	PTDA_ECOLI
11	48	33.1	601 1	PDML_DROME
12	48	33.1	750 1	MUTB_MICTU
13	47.5	32.8	229 1	CLCB_RAT
14	47.5	32.8	1509 1	GSRI_HUMAN
15	47	32.4	416 1	PAX6_CHICK
16	47	32.4	216 1	PAX6_CORJA
17	47	32.4	422 1	PAX6_HUMAN
18	47	32.4	422 1	PAX6_MOUSE
19	47	32.4	422 1	PAX6_XENLA
20	47	32.4	649 1	PTMA_VIBCH
21	47	32.4	1324 1	IRS2_HUMAN
22	46.5	32.1	162 1	PUR6_BACSU
23	46.5	32.1	419 1	CARP_CRYPA
24	46.5	32.1	1088 1	PIGO_HUMAN
25	46	31.7	208 1	FCP_ISOGA
26	46	31.7	396 1	G3PA_ARATH
27	46	31.7	442 1	AROA_BORBR
28	46	31.7	442 1	AROA_BORPE
29	46	31.7	748 1	PTIP_ECOLI
30	46	31.7	748 1	PTIP_ECOLI
31	46	31.7	952 1	UVRA_THIEH
32	46	31.7	1267 1	VL3_REOVD
33	46	31.7	1267 1	VL3_REOVD

34	46	31.7	1267 1	VL3_REOVL	P17376 reovirus (t
35	46	31.7	1434 1	VG65_HSV11	O00106 ictaluriid h
36	45.5	31.4	456 1	RINI_RAT	P29315 rattus norv
37	45.5	31.4	3801 1	LYST_HUMAN	O09698 homo sapien
38	45	31.0	111 1	RIA2_ASPEU	O90026 aspergillus
39	45	31.0	323 1	JUND_CHICK	P27921 gallus galli
40	45	31.0	341 1	VR15_MYCTU	O07314 mycobacteri
41	45	31.0	381 1	AMPC_ENTCL	P05364 enterobacte
42	45	31.0	578 1	VCSI_BACSU	P42967 bacillus su
43	45	31.0	730 1	EIS_HUMAN	P15502 homo sapien
44	45	31.0	2505 1	CCAA_HUMAN	O00555 homo sapien
45	45	31.0	4568 1	DYHB_CHLRE	O39565 chlamydomon

## ALIGNMENTS

```

RESULT 1
ID      7SBG_SOYBN      STANDARD:      PRT: 427 AA.
AC      P13917: Q39901; Q43464:
DT      01-JAN-1990 (Rel. 13, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Basic 7S globulin precursor (Bg) (Sbg7S).
GN      BG OR G7S.
OS      Glycine max (Soybean).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX      NCBI_TaxID=3847;
RN      [1]
RP      SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC      STRAIN=cv. Miyagishirume; TISSUE=Seed;
RX      MEDLINE=90067863; PubMed=2587227;
RA      Kagawa H., Hirano H.;
RT      "Sequence of a cDNA encoding soybean basic 7S globulin.";
RL      Nucleic Acids Res. 17:8868-8868(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Miyagishirume; TISSUE=Etolated leaf;
RX      MEDLINE=94336768; PubMed=8058830;
RA      Watanabe Y., Hirano H.;
RT      "Nucleotide sequence of the basic 7S globulin gene from soybean.";
RL      Plant Physiol. 105:1019-1020(1994).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Shi-shi; TISSUE=Cotyledon;
RA      Shu T.F., Hsieh K.L., Hsing Y.I., Chen Z.Y., Chow T.Y.;
RT      Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RN      [4]
RP      PARTIAL SEQUENCE.
RC      Kagawa H., Yamauchi F., Hirano H.;
RT      "Soybean basic 7S globulin represents a protein widely distributed in
RT      legume species.";
RL      FEBS Lett. 226:145-149(1987).
RN      [5]
RP      FUNCTION: SEED STORAGE PROTEIN. HAS A PROTEIN KINASE ACTIVITY.
RN      BINDS LEGNINOLIN.
RN      [6]
RP      FUNCTION: SEED STORAGE PROTEIN.
RN      [7]
RP      SUBUNIT: THE MATURE BG IS COMPOSED OF HIGH- AND LOW-KDa SUBUNITS
RN      WHICH ARE LINKED BY DISULFIDE BONDING.
RN      [8]
RP      This SWISS-PROT entry is copyright. It is produced through a collaboration
RN      between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN      the European Bioinformatics Institute. There are no restrictions on its
RN      use by non-profit institutions as long as its content is in no way
RN      modified and this statement is not removed. Usage by and for commercial
RN      entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN      or send an email to license@sib-sib.ch).
RN      [9]
RP      EMBL: X16469; CAA34489.1; -
RN      EMBL: U59425; AAB03390.1; -
RN      EMBL: D16107; BAA03681.1; -

```

DR PIR: S06750; S06750.  
 KW Glycoprotein; Seed storage protein; signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 427 BASIC 7S GLOBULIN.  
 FT CHAIN 25 427 HIGH-KDA SUBUNIT.  
 FT CHAIN 276 427 LOW-KDA SUBUNIT.  
 FT CONFLICT 48 48 W -> S (IN REF. 3).  
 FT CONFLICT 264 264 N -> T (IN REF. 1).  
 FT CONFLICT 305 305 F -> C (IN REF. 1).  
 FT CONFLICT 309 309 F -> C (IN REF. 1).  
 SQ SEQUENCE 427 AA; 46393 MW; 66041BC0680BACB CRC64;

Query Match 45.5%; Score 66; DB 1; Length 427;  
 Best Local Similarity 54.2%; Pred. No. 0.15;  
 Matches 13; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Oy 5 LPRSTGVAGLANSGLALPAOVAS 28  
 Db 175 LPRMTGVAGLGHAPISLPNOLAS 198  
 ||| : ||||| : : ||| : |||

RESULT 2  
 MYCN\_MARMO STANDARD; PRT: 460 AA.  
 ID MYCN\_MARMO  
 AC 261976;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE N-myc proto-oncogene protein (N-mycl).  
 GN MYCN OR NMYC OR MYCL.  
 OS Marmota monax (Woodchuck).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae;  
 OC Marmota.  
 CC NCBI\_TaxID=9995;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=90370481; PubMed=2395655;  
 RA Jorel G., Tjollals P., Buendia M.-A.;  
 RL Nucleotide sequence of the woodchuck N-myc gene (WN-mycl).";  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. BINDS DNA AS AN HETERODIMER WITH MAX.  
 CC -1- TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.  
 CC "this SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR IMBL: X53673; CA37712.1; -  
 DR IMBL: X53674; CA37712.1; JOINED.  
 DR HSSP: P25912; IHLO.  
 DR TRANSFAC: T03440; -  
 DR InterPro: IPR001092; HLH\_basic.  
 DR InterPro: IPR002418; TF\_Myc.  
 DR Pfam: PF00010; HLH\_1.  
 DR Pfam: PF01056; Myc\_N-term; 1.  
 DR PRINTS: PRO0044; LEUZIPPMYC.  
 DR SMART: SM00353; HLH\_1.  
 DR FROSITE: PS00038; HLH\_1; 1.  
 DR FROSITE: PS50888; HLH\_2; 1.  
 KW Nuclear protein; DNA-binding; Proto-oncogene; Phosphorylation.  
 FT LMAIN 258 274 ASP/GLU-RICH (ACIDIC).  
 FT LMAIN 377 390 BASIC DOMAIN.  
 FT LMAIN 391 430 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT LMAIN 429 450 LEUCINE-ZIPPER (POTENTIAL).  
 FT MOD\_RES 257 257 PHOSPHORYLATION (BY CK2)

FT FT (BY SIMILARITY).  
 FT MOD\_RES 259 259 PHOSPHORYLATION (BY CK2)  
 FT FT (BY SIMILARITY).  
 SQ SEQUENCE 460 AA; 49192 MW; 8A16686C825B02E CRC64;  
 SO SEQUENCE 11; Conservative 52.4%; Pred. No. 4.7;  
 Query Match 37.9%; Score 55; DB 1; Length 460;  
 Best Local Similarity 52.4%; Pred. No. 4.7;  
 Matches 11; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Oy 7 RSTGVAGLANSGLALPAOVA 27  
 Db 159 RGHGTAGAGAGALPAELA 179  
 ||| : ||| : : ||| : |||

RESULT 3  
 MYCN\_HUMAN STANDARD; PRT: 464 AA.  
 ID MYCN\_HUMAN  
 AC P04198;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE N-myc proto-oncogene protein.  
 GN MYCN OR NMYC.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=66149381; PubMed=2869488;  
 RA Stanton L.W., Schwab M., Bishop J.M.;  
 RL Nucleotide sequence of the human N-myc gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1772-1776(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86092232; PubMed=3510398;  
 RA Kohl N.E., Legouy E., Depinho R.A., Nisen P.D., Smith R.K.,  
 RA Gee C.E., Alt F.W.;  
 RL "Human N-myc is closely related in organization and nucleotide  
 RL sequence to c-myc.";  
 RL Nature 319:73-77(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88202932; PubMed=2834684;  
 RA Ibsen J.M., Rablitts P.H.;  
 RL "Sequence of a germ-line N-myc gene and amplification as a mechanism  
 RL of activation.";  
 RL Oncogene 2:399-402(1988).  
 RN [4]  
 RP SEQUENCE OF 71-464 FROM N.A.  
 RX MEDLINE=86179901; PubMed=3008339;  
 RA Slamon D.J., Boone T.C., Seeger R.C., Keith D.E., Chazin V., Lee H.C.,  
 RA Souza L.M.;  
 RL "Identification and characterization of the protein encoded by the  
 RL human N-myc oncogene.";  
 RL Science 232:768-772(1986).  
 RN [5]  
 RP SEQUENCE OF 313-464 FROM N.A.  
 RX MEDLINE=85215633; PubMed=2987858;  
 RA Michitsch R.W., Melera P.W.;  
 RL "Nucleotide sequence of the 3' exon of the human N-myc gene.";  
 RL Nucleic Acids Res. 13:2545-2558(1985).  
 RN [6]  
 RP PHOSPHORYLATION BY CK2.  
 RX MEDLINE=93049352; PubMed=1425701;  
 RA Hagihara T., Nakaya K., Nakamura Y., Nakajima H., Nishimura S.,  
 RA Taya Y.;  
 RL "Specific phosphorylation of the acidic central region of the N-myc  
 RL protein by casein kinase II.";  
 RL Eur. J. Biochem. 209:945-950(1992).  
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. BINDS DNA AS AN HETERODIMER WITH MAX.

```

CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING FETAL DEVELOPMENT.
CC -1- DISEASE: AMPLIFICATION OF THE N-MYC GENE IS ASSOCIATED WITH A
CC VARIETY OF HUMAN TUMORS, MOST FREQUENTLY NEUROBLASTOMA, WHERE THE
CC LEVEL OF AMPLIFICATION APPEARS TO INCREASE AS THE TUMOR
CC PROGRESSES.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chromancer/Genes/NMYC112.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03294; CAA27037.1; -.
DR EMBL: X03295; CAA27038.1; -.
DR EMBL: M13241; AAA36371.1; ALT_INIT.
DR EMBL: M13228; AAA36370.1; -.
DR EMBL: X02363; -. NOT_ANNOTATED_CDS.
DR EMBL: Y00664; CAA68678.1; ALT_INIT.
DR PIR: A25744; TVHUM2.
DR PIR: A01355; TVHUMC.
DR PIR: A22937; A22937.
DR PIR: S02249; S02249.
DR HSSP: P25912; IHLO.
DR TRANSFAC: T02379; -.
DR Genew: HGNC:7559; MYCN.
DR MIM: 164840; -.
DR InterPro: IPR001092; HLH_basic.
DR InterPro: IPR002418; TF_Myc.
DR Pfam: PF00010; HLH_1.
DR Pfam: PF01056; MYC_N_term; 1.
DR PRINTS: PR00044; LEUZIPPRMYC.
DR SMART: SM00353; HLH_1.
DR PROSITE: PS00038; HLH_1; 1.
DR PROSITE: PS50888; HLH_2; 1.
KW Nuclear protein; DNA-binding; Proto-oncogene; Phosphorylation.
FT DOMAIN 262 278 ASP/GLU-RICH (ACIDIC).
FT DNA_BIND 381 394 BASIC DOMAIN.
FT DOMAIN 395 434 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 433 454 LEUCINE-ZIPPER (POTENTIAL).
FT MOD_RES 261 261 PHOSPHORYLATION (BY CK2).
FT MOD_RES 263 263 PHOSPHORYLATION (BY CK2).
FT CONFLICT 227 227 A -> P (IN REF. 2).
FT CONFLICT 363 363 I -> V (IN REF. 3).
SQ SEQUENCE 464 AA; 49561 MW; 560E885602E30DAD CRC64;

```

```

Query Match 37.9%; Score 55; DB 1; Length 464;
Best Local Similarity 52.4%; Pred. No. 4.8;
Matches 11; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 7 RSTGVAGLANSGLAPAOVA 27
DB 160 RGHGAGAGAGAGALPAEIA 180

```

```

RESULT 4
DIF1_CAEEL STANDARD; PRT; 312 AA.
AC 027257; Q20630;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein dif-1.
DE DIF-1 OR F49E8.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.

```

```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RX MEDLINE-95292971; PubMed-7774589;
RA Ahinger J.;
RT "Embryonic tissue differentiation in Caenorhabditis elegans requires
RT dif-1, a gene homologous to mitochondrial solute carriers.";
RL EMBO J. 14:2307-2316(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-95102108; PubMed-7803813;
RA Runswick M.J., Philippides A., Lauria G., Walker J.E.;
RT "Extension of the mitochondrial transporter super-family: sequences of
RT five members from the nematode worm, Caenorhabditis elegans.";
RL DNA Seq. 4:281-291(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Murray J., Wohlmann P.;
RN Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RP REVISIONS.
RA Waterston R.;
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: Seems to play a role in the maintenance of tissue
CC differentiation in the developing embryo, but not for its
CC initiation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (potential).
CC -1- DEVELOPMENTAL STAGE: Although dif-1 activity is required only
CC during embryogenesis, it is expressed at all stages of
CC development.
CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z48240; CAA88283.1; -.
DR EMBL: X76115; CAA53721.1; -.
DR EMBL: U61949; AAB03153.2; -.
DR Wormep: F49E8.5; CE28408.
DR InterPro: IPR001993; Mitoch_carrier.
DR Pfam: PF00153; mito_car; 3.
DR PROSITE: PS00215; MITOCH_CARRIER; 3.
KW Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport;
KW Developmental protein.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 209 229 POTENTIAL.
FT TRANSMEM 277 297 POTENTIAL.
SQ SEQUENCE 312 AA; 33134 MW; CCF1300CAE4D64F CRC64;

```

```

Query Match 36.6%; Score 53; DB 1; Length 312;
Best Local Similarity 38.7%; Pred. No. 6;
Matches 12; Conservative 8; Mismatches 7; Indels 4; Gaps 1;

```

```

OY 4 SLPRGST---GVAGLANSGLAPAOVASQ 30
DB 202 TLSPGATLWAGLGIANNVCIPADVLSR 232

```

```

RESULT 5
YZ08_METJA STANDARD; PRT; 520 AA.
AC 060270;
DT 01-NOV-1997 (rel. 35, Created)

```



AC O15119; G9UKF8;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE T-box transcription factor TBX3 (T-box protein 3).  
 GN TBX3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM I), AND FUNCTION.  
 RC TISSUE-Breast carcinoma;  
 RX MEDLINE=99398688; PubMed=10468588;  
 RA He M.-L., Wen L., Campbell C.E., Wu J.Y., Rao Y.;  
 RT "Transcription repression by Xenopus ET and its human ortholog TBX3, a  
 gene involved in ulnar-mammary syndrome.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:10212-10217(1999).  
 RN [2]  
 RP SEQUENCE OF 1-488 FROM N.A. (ISOFORM I).  
 RC TISSUE-Kidney;  
 RX MEDLINE=97351519; PubMed=9207801;  
 RA Banashid M., Lin R.C., Law D.J., Watkins W.S., Krakowiak P.A.,  
 RA Moore M.E., Franceschini P., Lala R., Holmes L.B., Gebuhr T.C.,  
 RA Schinzel A., Bruneau B.G., Seidman J.G., Seidman C.E., Jorde L.B.;  
 RT "Mutations in human TBX3 alter limb, apocrine and genital development  
 in ulnar-mammary syndrome.";  
 RL Nat. Genet. 16:311-315(1997).  
 RN [3]  
 RP SEQUENCE OF 591-742 FROM N.A., ALTERNATIVE SPLICING, AND VARIANTS UMS.  
 RX MEDLINE=99264236; PubMed=10330342;  
 RA Banashid M., Le T., Watkins W.S., Dixon M.E., Kramer B.E., Roeder A.D.,  
 RA Carey J.C., Root S., Schinzel A., Van Maldergem L., Gardner R.J.M.,  
 RA Lin R.C., Seidman C.E., Seidman J.G., Wallenstein R., Moran E.,  
 RA Stupphen R., Campbell C.E., Jorde L.B.;  
 RT "The spectrum of mutations in TBX3: genotype/phenotype relationship in  
 ulnar-mammary syndrome.";  
 RL Am. J. Hum. Genet. 64:1550-1562(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM III).  
 RC TISSUE-Adrenal gland;  
 RA Song H., Gao G., Peng Y., Ren S., Chen Z., Han Z.;  
 RT "A novel gene expressed in human adrenal gland.";  
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR INVOLVED IN DEVELOPMENTAL  
 PROCESSES. PROBABLY PLAYS A ROLE IN LIMB PATTERN FORMATION.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS: I, II (SHOWN HERE) AND  
 III; ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORMS II AND III  
 CONTAIN AN INTERRUPTED T-BOX DOMAIN. AN ADDITIONAL ISOFORM IV MAY  
 BE PRODUCED BY JOINING EXON 1 TO EXON 7 THEREBY ELIMINATING THE T-  
 BOX.  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.  
 CC -1- DISEASE: DEFECTS IN TBX3 ARE THE CAUSE OF ULNAR-MAMMARY SYNDROME  
 (UMS). THIS DISEASE IS CHARACTERIZED BY ULNAR RAY DEFECTS,  
 OBESITY, HYPOGENITALISM, DELAYED PUBERTY, HYPOPLASIA OF NIPPLES  
 AND APOCRINE GLANDS.  
 CC -1- SIMILARITY: CONTAINS 1 T-BOX DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: AF170708; AAD50989.2; -  
 CC EMBL: AF002228; AAC12947.1; -  
 CC EMBL: AF140240; AAF61816.1; -  
 CC EMBL: AF216750; AAF61207.1; -  
 CC HSSP: P24781; 1XBR.  
 CC TRANSFAC: T04413; -

DR TRANSFAC: T04414; -  
 DR Genew: HGNC:11602; TBX3.  
 DR MIM: 601621; -  
 DR MIM: 181450; -  
 DR InterPro: IPR001699; TF\_T-box.  
 DR Pfam: PF00907; T-box; 1.  
 DR PRINTS: PR00937; TBOX.  
 DR SMART: SM00425; TBOX\_1.  
 DR PROSITE: PS01283; TBOX\_1; 1.  
 DR PROSITE: PS01264; TBOX\_2; 1.  
 DR PROSITE: PS0252; TBOX\_3; 1.  
 DR Transcription regulation; DNA-binding; Repressor; Nuclear protein;  
 KW developmental protein; Disease mutation; Alternative splicing.  
 FT DNA\_BIND 112 220  
 FT DNAS\_BIND 241 304  
 FT DOMAIN 544 694  
 FT VARSPLIC 221 240  
 FT VARSPLIC 490 628  
 FT  
 FT VANSPLIC 660 676  
 FT VARIANT 143 143  
 FT  
 FT VARIANT 149 149  
 FT  
 FT CONFLICT 315 315  
 FT CONFLICT 616 627  
 FT  
 FT CONFLICT 673 673  
 FT CONFLICT 691 691  
 FT SEQUENCE 742 AA: 79402 MW: D2178A2480962160 CRC64;  
 QY 1 LLASLRPGSTGVAGLANSGLAPAVASMQ 30  
 DB 542 LLAIVSGASTGVSGLDSTAM--ASAAAQ 568  
 Query Match 34.1%; Score 49.5; DB 1; Length 742;  
 Best Local Similarity 43.3%; Pred. No. 40;  
 Matches 13; Conservative 7; Mismatches 7; Indels 3; Gaps 1;  
 RESULT 8  
 SPEH\_METUA STANDARD: PRT: 124 AA.  
 ID SPEH\_METUA  
 AC Q57763;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50) (AdometDC)  
 DE (Samdc) [contains: S-adenosylmethionine decarboxylase beta chain; S-  
 DE adenosylmethionine decarboxylase alpha chain].  
 GN SPEH OR M20315.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 CC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Sult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kervatave A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073(1996).

[2]  
 RN CHARACTERIZATION.  
 RP STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RC MEDLINE-20528316; PubMed-11073910;  
 RX Kim A.D., Graham D.E., Seeholzer S.H., Markham G.D.;  
 RA "5-adenosylmethionine decarboxylase from the archaeon Methanococcus  
 RT jannaschii: identification of a novel family of pyruvoyl enzymes.";  
 RL J. Bacteriol. 182:6667-6672(2000).  
 CC -1- FUNCTION: DECARBOXYLATION OF S-ADENOSYLMETHIONINE PROVIDES THE  
 CC AMINOPROPYL MOIETY REQUIRED FOR SPERMIDINE BIOSYNTHESIS FROM  
 CC PUTRESCINE.  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine = (S)-deoxy-5-  
 CC adenosyl(1(3-aminopropyl) methylsulfonium salt + CO(2).  
 CC -1- COFACTOR: Pyruvoyl group.  
 CC -1- SUBUNIT: HETEROTETRAMER OF TWO ALPHA AND TWO BETA CHAINS.  
 CC -1- MASS SPECTROMETRY: MM=6794.5; METHOD=MALDI; RANGE=1-63.  
 CC -1- MASS SPECTROMETRY: MM=6991.6; METHOD=MALDI; RANGE=64-124.  
 CC -1- MISCELLANEOUS: THERMOSTABLE.  
 CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC ADOMETDC FAMILY. SUBFAMILY  
 CC 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U67486; AAB98301.1; ALT\_INIT.  
 DR TIGR: MJ0315; -  
 DR InterPro: IPR003826; SAMDC.  
 DR Pfam: PF02675; DUF206; 1.  
 KW Spermidine biosynthesis; Lyase; Decarboxylase; zymogen; Pyruvate;  
 KW Complete proteome.  
 FT CHAIN 1 63 S-ADENOSYLMETHIONINE DECARBOXYLASE BETA  
 FT CHAIN 64 124 CHAIN.  
 FT CHAIN S-ADENOSYLMETHIONINE DECARBOXYLASE ALPHA  
 FT CHAIN.  
 FT SITE 63 64 CLEAVAGE (NONHYDROLYTIC).  
 FT MOD\_RES 64 64 CONVERTED TO A PYRUVOL GROUP.  
 FT ACT\_SITE 84 84 IMPORTANT FOR CATALYTIC ACTIVITY (BY  
 FT SIMILARITY).  
 FT SEQUENCE 124 AA: 13792 MW: BC1152CDE080F113 CRC64;  
 SO  
 Query Match 33.8%; Score 49; DB 1; Length 124;  
 Best Local Similarity 58.8%; Pred. No. 8.2;  
 Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

RT nitrous-oxide reductase from Paracoccus denitrificans. New and  
 RT conserved structural and regulatory motifs."  
 RL Eur. J. Biochem. 218:49-57(1993).  
 CC -1- FUNCTION: NITROUS-OXIDE REDUCTASE IS PART OF A BACTERIAL  
 CC RESPIRATORY SYSTEM WHICH IS ACTIVATED UNDER ANAEROBIC  
 CC CONDITIONS IN THE PRESENCE OF NITRATE OR NITROUS OXIDE.  
 CC -1- CATALYTIC ACTIVITY: N(2) + H(2)O + acceptor = nitrous oxide +  
 CC reduced acceptor.  
 CC -1- COFACTOR: COPPER (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER; EACH SUBUNIT CONTAINS 2 DINUCLEAR COPPER  
 CC CENTERS A AND 2. 2 IS THOUGHT TO BE THE SITE OF NITROUS OXIDE  
 CC REDUCTION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, TO MITOCHONDRIAL AND  
 CC BACTERIAL COX2 SUBUNITS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X74792; CA52798.1; -  
 KW Signal; Oxidoreductase; Copper; Metal-binding; Periplasmic.  
 FT SIGNAL 1 57 BY SIMILARITY.  
 FT CHAIN 58 652 NITROUS-OXIDE REDUCTASE.  
 FT METAL 595 595 COPPER A 1 (BY SIMILARITY).  
 FT METAL 630 630 COPPER A 1 AND 2 (BY SIMILARITY).  
 FT METAL 632 632 COPPER A 2 (BY SIMILARITY).  
 FT METAL 634 634 COPPER A 1 AND 2 (BY SIMILARITY).  
 FT METAL 638 638 COPPER A 2 (BY SIMILARITY).  
 FT METAL 641 641 COPPER A 1 (BY SIMILARITY).  
 FT METAL 641 641 COPPER A 1 (BY SIMILARITY).  
 SQ SEQUENCE 652 AA: 71413 MW: 40492A4FDE7EDEA8 CRC64;  
 SO  
 Query Match 33.8%; Score 49; DB 1; Length 652;  
 Best Local Similarity 54.5%; Pred. No. 41;  
 Matches 12; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 ASLPRGSTGVAGLANGLALPA 24  
 DB 37 AALGLGTAGVATVAGSAAALAA 58

RESULT 10  
 PRDA\_ECOLI STANDARD: PRT: 485 AA.  
 ID PTDA\_ECOLI  
 AC P24241; Q46880;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE PTS system, arbutin-, cellobiose-, and salicin-specific IIABC  
 DE component (EIIBC-ASC) (Arbutin-, cellobiose-, and salicin-permease  
 DE IIABC component) (Phosphotransferase enzyme II, ABC component)  
 DE (EC 2.7.1.69) (EIIBC-ASC).  
 GN ASCF OR B2713.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-92334140; PubMed-1630307;  
 RA Hall B.G., Xu L.;  
 RT "Nucleotide sequence, function, activation, and evolution of the  
 RT cryptic asc operon of Escherichia coli K12.";  
 RL Mol. Biol. Evol. 9:688-706(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE-97426617; PubMed-9278503;

RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
CC -I- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE  
CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY  
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HRP). IIA TRANSFERS ITS  
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
CC THE SUGAR.  
CC -I- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein  
CC histidine + sugar phosphate.  
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
CC -I- SIMILARITY: CONTAINS 1 PTS EIIA DOMAIN.  
CC -I- SIMILARITY: CONTAINS 1 PTS EIIIB DOMAIN.  
CC -I- SIMILARITY: CONTAINS 1 PTS EIIC DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M73326; AAA16429.1; -  
CC EMBL: U29579; AAA69225.1; -  
CC EMBL: AE000355; AAC75757.1; -  
CC PIR: B44070; B44070.  
CC PIR: S27552; S27552.  
CC HSSP: P05053; 1IBA.  
CC Ecogene: EGI0086; ascf.  
CC InterPro: IPR004719; PTSIIC-glic.  
CC InterPro: IPR001996; PTS-EIIB.  
CC InterPro: IPR003352; PTS-EIIC.  
CC Pfam: PF00367; PTS-EIIB.1.  
CC Pfam: PF02378; PTS-EIIC.1.  
CC Pfam: PF001476; PTS-EIIB.1.  
CC Pfam: PF001476; PTS-EIIB.1.  
CC TIGRfams: TIGR00826; EIIB.1.  
CC TIGRfams: TIGR00852; ptc-glic.1.  
CC PROSITE: PS01035; PTS-EIIB\_CVS.1.  
CC KW Phosphotransferase system; Sugar transport; Transferase;  
CC Phosphorylation; Transmembrane; Inner membrane; Complete proteome.  
CC FT DOMAIN 1 43 EIIB DOMAIN.  
CC FT DOMAIN ? ? EIIC DOMAIN.  
CC FT MOD\_RES ? 485 EIIC DOMAIN.  
CC FT MOD\_RES 28 28 PHOSPHORYLATION (BY SIMILARITY).  
CC FT TRANSMEM 102 122 PHOSPHORYLATION (BY SIMILARITY).  
CC FT TRANSMEM 146 166 POTENTIAL.  
CC FT TRANSMEM 177 197 POTENTIAL.  
CC FT TRANSMEM 207 227 POTENTIAL.  
CC FT TRANSMEM 254 274 POTENTIAL.  
CC FT TRANSMEM 285 305 POTENTIAL.  
CC FT TRANSMEM 330 350 POTENTIAL.  
CC FT TRANSMEM 363 383 POTENTIAL.  
CC FT TRANSMEM 389 409 POTENTIAL.  
CC FT TRANSMEM 433 453 POTENTIAL.  
CC FT CONFLICT 167 170 HLPR -> OSAA (IN REF. 1).  
CC FT CONFLICT 311 311 R -> H (IN REF. 1).  
CC SEQUENCE 485 AA; 51229 MW; CFEEDDD8CABBC38 CMC64;  
CC  
CC Query Match 33.4%; Score 48.5; DB 1; Length 485;  
CC Best Local Similarity 40.0%; Pred. No. 36;  
CC Matches 14; Conservative 4; Mismatches 12; Indels 5; Gaps 1;

RESULT 11  
ID PDM1\_DROME STANDARD: PRT: 601 AA.  
AC P31368;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Nubbin protein (Twain protein) (POU domain protein 1) (PDM-1) (DPOU-  
DE 19) (DPOU1).  
GN NUB OR TWN OR PDM-1 OR POU-19 OR OCT1.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92001544; PubMed=1680380;  
RA Billin A.N., Cockerill K.A., Poole S.J.;  
RT "Isolation of a family of Drosophila POU domain genes expressed in  
RT early development.";  
RL Mech. Dev. 34:75-84(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92144419; PubMed=1685891;  
RA Lloyd A., Sakonju S.;  
RT "Characterization of two Drosophila POU domain genes, related to  
RT oct-1 and oct-2, and the regulation of their expression patterns.";  
RL Mech. Dev. 36:87-102(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91352045; PubMed=1881906;  
RA Dick T., Yang X., Yeo S., Chia W.;  
RT "Two closely linked Drosophila POU domain genes are expressed in  
RT neuroblasts and sensory elements.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:7645-7649(1991).  
CC -I- FUNCTION: DNA-BINDING REGULATORY PROTEIN IMPLICATED IN EARLY  
CC DEVELOPMENT. INVOLVED IN NEURONAL CELL FATE DECISION. REPRESSED  
CC DIRECTLY OR INDIRECTLY BY THE BX-C HOMEOLOGIC PROTEINS.  
CC -I- SUBCELLULAR LOCATION: Nuclear.  
CC -I- TISSUE SPECIFICITY: NEUROBLASTS AND SENSORY ELEMENTS.  
CC -I- DEVELOPMENTAL STAGE: EXPRESSED PRIMARILY DURING THE FIRST HALF OF  
CC EMBRYOGENESIS. INITIAL EXPRESSION IN CELLULAR BLASTODERM STAGE,  
CC THEN IN ECTODERMAL STRIPES DURING GERMBAND EXTENSION. BROAD  
CC EXPRESSION IN THE NEUROECTODERM FOLLOWED BY LIMITATION TO DISCRETE  
CC SUBSETS OF CNS CELLS, AND EXPRESSION IN SPECIFIC PNS NEURONS AND  
CC SUPPORT CELLS.  
CC -I- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.  
CC -I- SIMILARITY: CONTAINS 1 HOMEOBOX DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M81957; AAA28829.1; -  
CC EMBL: S80561; AAR21409.1; -  
CC EMBL: M65015; AAA28480.1; -  
CC PIR: A41277; A41277.  
CC HSSP: P14859; 1OCT.  
CC TRANSFAC: T01900; -  
CC FLYbase: FBgn0002970; nub.  
CC InterPro: IPR001356; Homeobox.  
CC InterPro: IPR00327; POU\_domain.  
CC Pfam: PF00046; homeobox.1.  
CC Pfam: PF00157; pou.1.  
CC PRINTS: PR00028; POUDOMAIN.

```

DR ProDom: PD000010; Homeobox; 1.
DR ProDom: PD000583; POU domain; 1.
DR SMART: SM00389; HOX; 1.
DR SMART: SM00352; POU; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00035; POU_1; 1.
DR PROSITE: PS00465; POU_2; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Transcription regulation; Nuclear protein;
KW Developmental protein;
FT DOMAIN 104 107 POLY-HIS.
FT DOMAIN 191 201 POLY-ALA.
FT DOMAIN 359 364 POLY-HIS.
FT DOMAIN 407 415 POLY-ALA.
FT DOMAIN 425 495 POU.
FT DNA_BIND 523 582 HOMEBOX.
FT CONFLICT 420 420 P->R (IN REF. 2).
SQ SEQUENCE 601 AA; 65202 MW; 8FFCD7B3C162D2B2 CRC64;

Query Match 33.1%; Score 48; DB 1; Length 601;
Best Local Similarity 57.1%; Pred. No. 52;
Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 10 TGVAGLANSGLALPAQVAAQ 30
Db 272 SGLSLADDDPALTAQVAAQ 292

RESULT 12
MUTB_MYCTU STANDARD: PRT; 750 AA.
AC P71774;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable methylmalonyl-CoA mutase large subunit (EC 5.4.99.2) (MCM).
GN MUTB OR RV1493 OR MT1540 OR MTCY277.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Jordan S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badoek K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence."
RT Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-CDC 1551 / Oshkosh;
RC Jeltschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Jeltschmann R., Debey R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains."
RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE ISOMERIZATION OF SUCCINYL-CoA TO
CC METHYLMALONYL-CoA DURING SYNTHESIS OF PROPIONATE FROM
CC TRICARBOXYLIC ACID-CYCLE INTERMEDIATES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: (R)-2-methyl-3-oxopropanoyl-CoA -> succinyl-
CC CoA.
CC -1- COFACTOR: ADENOSYLCOBALAMIN (BY SIMILARITY).
```

```

CC -1- PATHWAY: Propionic acid fermentation.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE METHYLMALONYL-CoA MUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z79701; CAB02043.1; -
DR EMBL: AE007022; AK45807.1; -
DR HSSP: P11653; IREQ.
DR TIGR: MT1540; -
DR Tuberculist; RV1493; -
DR InterPro: IPR003312; B12-binding.
DR InterPro: IPR001629; MM_COA_mutase.
DR Pfam: PF01642; MM_COA_mutase; 1.
DR Pfam: PF02310; B12-binding; 1.
DR TIGRFAMs: TIGR00640; acid_COA_mut_C; 1.
DR TIGRFAMs: TIGR00641; acid_COA_mut_N; 1.
DR PROSITE: PS00544; METHYLMALONYL-CoA_MUTASE; 1.
KM Isomerase; Vitamin B12; Cobalt; Complete proteome.
FT DOMAIN 587 626 METHYLMALONYL-CoA-BINDING SITE (BY
FT SIMILARITY).
FT METAL 629 629 COBALAMIN-BINDING (POTENTIAL).
FT FT 629 629 COBALT.
SQ SEQUENCE 750 AA; 80604 MW; 9E6B86CE507D022 CRC64;

OY 3 ASLPRGTVAGLANSGLALPAQVAAQ 30
Db 534 AAEGGRAGADGNNLALALDARAQ 561

RESULT 13
CLCB_RAT STANDARD: PRT; 229 AA.
AC P08082;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Clathrin light chain B (Lcb).
GN CLTB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87178007; PubMed=3563513;
RA Kirchhausen T., Scarlato P., Harrison S.C., Monroe J.J., Chow E.P.,
RA Mattaliano R.J., Ramachandran K.L., Smart J.E., Ahn A.H., Brosius J.;
RA "Clathrin light chains LCA and LCB are similar, polymorphic, and
RA share repeated heptad motifs."
RT Science 236:320-324(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA "FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
RA COATED PITS AND VESICLES.
RA SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3
RA HEAVY CHAINS AND 3 LIGHT CHAINS.
RA -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
RA VESICLES.
RA -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, BRAIN (SHOWN HERE) AND NON-
RA BRAIN; ARE PRODUCED BY ALTERNATIVE SPLICING.
RA -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```



CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M15883; AAA40880.1; -  
 DR EMBL; M19262; AAA40891.1; -  
 DR PIR; B25994; LRRTB2.  
 DR InterPro; IPR000996; Clathrin\_19\_ch.  
 DR Pfam; PF01086; Clathrin\_19\_ch; 1.  
 DR PROSITE; PS00024; CLATHRIN\_LIGHT\_CHAIN\_2; 1.  
 DR PROSITE; PS00581; CLATHRIN\_LIGHT\_CHAIN\_2; 1.  
 DR Coated pits; Alternative splicing; Calcium-binding; Phosphorylation.  
 KW MOD\_RES 11 11 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 13 13 PHOSPHORYLATION (BY SIMILARITY).  
 FT DOMAIN 93 155 INVOLVED IN BINDING CLATHRIN HEAVY  
 FT CHAIN.  
 FT DISULFID 199 209 BY SIMILARITY.  
 FT VARSPLIC 156 173 MISSING (IN ISOFORM NON-BRAIN).  
 SQ SEQUENCE 229 AA; 25117 MW; C4A05B07CEB92A85 CRC64;  
 Query Match 32.8%; Score 47.5; DB 1; Length 229;  
 Best Local Similarity 48.4%; Pred. No. 23;  
 Matches 15; Conservative 4; Mismatches 9; Indels 3; Gaps 2;  
 OY 3 ASLPRGSTGVAGLAN-SGLALPA--OVASAO 30  
 DB 26 AFLAQOESFIAGIENDSGFCAPARASOVASAO 56  
 RESULT 14  
 GSRL\_HUMAN  
 ID GSRL\_HUMAN STANDARD; PRT; 1509 AA.  
 AC OGNZMA:  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 16-OCT-2001 (Rel. 40; Last annotation update)  
 DE Glioma tumor suppressor candidate region gene 1 protein.  
 GN GLTSCR1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.  
 RX MEDLINE=20175430; PubMed=10708517;  
 RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,  
 RA Portier B.P., Ueki K., Billings S., Ramasamy S., Mohrenweiser H.W.,  
 RA Scheithauer B.W., Louis D.N., Jenkins R.B.;  
 RT "A transcript map of the chromosome 19q-Arm glioma tumor suppressor  
 RT region."  
 RL Genomics 64:44-50(2000).  
 CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,  
 CC placenta, skeletal muscle, and pancreas, and at lower levels in  
 CC lung, liver, and kidney.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF182077; AAF62874.1; -  
 DR GenBank; HGNC:4332; GLTSCR1.  
 DR MIM; 605690; -  
 FT DOMAIN 37 45 POLY-GLY.  
 FT DOMAIN 884 889 POLY-PRO.  
 FT DOMAIN 1214 1225 POLY-SER.  
 FT DOMAIN 1282 1286 POLY-PRO.  
 FT DOMAIN 1284 1304 POLY-PRO.  
 SQ SEQUENCE 1509 AA; 152991 MW; 7C5144F443CE6821 CRC64;

Query Match 32.8%; Score 47.5; DB 1; Length 1509;  
 Best Local Similarity 56.0%; Pred. No. 1,5e+02;  
 Matches 14; Conservative 0; Mismatches 10; Indels 1; Gaps 1;  
 OY 2 LASLPRGSTGVAGLAN-SGLALPAOV 26  
 DB 163 LOGLPNGSPGATPAATGLA-PIOV 186  
 RESULT 15  
 PAX6\_CHICK  
 ID PAX6\_CHICK STANDARD; PRT; 216 AA.  
 AC P47237;  
 DT 01-FEB-1996 (Rel. 33; Created)  
 DT 01-FEB-1996 (Rel. 33; Last sequence update)  
 DT 15-JUN-2002 (Rel. 41; Last annotation update)  
 DE Paired box protein Pax-6 (Fragment).  
 GN PAX-6.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94170909; PubMed=8125186;  
 RA Li H.S., Yang J.M., Jacobson R.D., Pasko D., Sundin O.;  
 RT "Pax-6 is first expressed in a region of ectoderm anterior to the  
 RT early neural plate: implications for stepwise determination of the  
 RT lens."  
 RL Dev. Biol. 162:181-194(1994).  
 CC -!- FUNCTION: MAY BE A TRANSCRIPTION FACTOR WITH IMPORTANT FUNCTIONS  
 CC IN EYE AND NASAL DEVELOPMENT.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; S69508; AAB30163.1; -  
 DR HSSP; P06601; 1FJL.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox; 1.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; 1.  
 DR PROSITE; PS50071; HOMEBOX\_2; 1.  
 KW Transcription regulation; Homeobox; DNA-binding; Nuclear protein;  
 KW Developmental protein.  
 FT NON\_TER 1 1  
 FT DNA\_BIND 10 69 HOMEBOX.  
 FT DOMAIN 79 216 PRO/SER/THR-RICH.  
 SQ SEQUENCE 216 AA; 24469 MW; 08CE98AD1F58333A CRC64;  
 Query Match 32.4%; Score 47; DB 1; Length 216;  
 Best Local Similarity 34.5%; Pred. No. 26;  
 Matches 10; Conservative 8; Mismatches 9; Indels 2; Gaps 1;  
 OY 4 SLPRGSTGV--AGLAN-SGLALPAQVASAO 30  
 DB 177 SCPMGTSTTSTGLISPGVSPVQVPGSE 205  
 Search completed: May 20, 2003, 16:35:42  
 Job time : 11.3462 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:19:09 ; Search time 38.3242 Seconds  
(without alignments)  
166.669 Million cell updates/sec

Title: US-09-869-155-16  
Perfect score: 145  
Sequence: 1 LASLPRGSTGVAGLANGLALPAQVASAQK 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	83.4	419	10	O8S1V0 oryza sativ
2	94	64.8	424	10	O8S1V1 oryza sativ
3	93	64.1	422	10	O8S1V3 oryza sativ
4	88	60.7	434	10	O8S1V5 oryza sativ
5	84	57.9	434	10	O8S1U4 oryza sativ
6	84	57.9	442	10	O8S1U6 oryza sativ
7	68	46.9	433	10	O05929 daucus caro
8	64	44.1	433	10	O9ZV54 daucus caro
9	64	44.1	433	10	O8RVH5 glycine max
10	63	43.4	369	10	O9FS29 glycine max
11	63	43.4	402	10	O8S1U9 oryza sativ
12	62	42.8	434	10	O9ZV55 arabidopsis
13	59	40.7	386	10	O9ZV53 arabidopsis
14	57.5	39.7	443	10	O8S1U7 oryza sativ
15	53.5	36.9	411	16	O9ZSC3 rhizobium m
16	53	36.6	416	16	O8XR46 ralstonia s

17	52	35.9	513	3	O9Y7P3 schizosach
18	51	35.2	231	2	O93NE1 myxococcus
19	51	35.2	825	2	O9Z6O5 rhodobacter
20	50.5	34.8	503	16	O9R6G4 delinococcus
21	50.5	34.8	1093	3	O8X0R0 neuropept
22	50	34.5	583	10	O946Y6 chlamydom
23	49.5	34.1	310	2	O560O8 streptomyce
24	49.5	34.1	653	10	O8W5H2 oryza sativ
25	49.5	34.1	672	16	O91OK3 pseudomonas
26	49.5	34.1	743	4	O8TB20 homo sapien
27	49.5	34.1	1443	16	O9RUC7 delinococcus
28	49.5	33.8	1752	5	O07265 strongyloce
29	49	33.8	210	2	O9RNU2 streptococc
30	49	33.8	217	1	O8X276 halobacteri
31	49	33.8	290	5	O9YTS1 dirosophila
32	49	33.8	377	2	O9X5C0 escherichia
33	49	33.8	377	2	O9WVU6 escherichia
34	49	33.8	377	16	O8XDQ2 escherichia
35	49	33.8	653	16	O9RDB5 streptomyce
36	49	33.8	1010	2	O9XCV5 cellulomona
37	49	33.8	2237	5	O9YI22 dirosophila
38	48.5	33.4	162	16	O8UEY4 agrobacteri
39	48.5	33.4	166	16	O9CN27 pasteurilla
40	48.5	33.4	485	16	O8X842 escherichia
41	48.5	33.4	1222	2	O9X349 bacillus an
42	48	33.1	189	16	O8YGI7 brucella me
43	48	33.1	208	2	O87372 acetobacter
44	48	33.1	244	16	O53721 mycobacteri
45	48	33.1	279	16	O987A8 rhizobium 1

## ALIGNMENTS

RESULT 1	O8S1V0	PRELIMINARY:	PRT:	419 AA.
AC	O8S1V0:			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Putative dermal glycoprotein.			
GN	P0504E02.9.			
OS-	Oryza sativa (japonica cultivar-group).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	Ehrhartoideae; Oryzeae; Oryza.			
OX	NCBI_TaxID=39947;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=CV. NIPPONBARE;			
RA	Sasaki T., Matsumoto T., Yamamoto K.;			
RT	"Oryza sativa nipponbare(GAS) genomic DNA, chromosome 1, PAC			
RT	clone:P0504E02."			
RL	Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: AP003269; BAB89708.1; -			
SD	SEQUENCE 419 AA: 43786 MW; 7928E5F5AF3F3A8 CRC64;			
Query Match	83.4%; Score 121; DB 10; Length 419;			
Best local similarity	86.7%; Pred. No. 3.1e-08;			
Matches	26; Conservative	2; Mismatches	2; Indels	0; Gaps
OY	1 LASLPRGSTGVAGLANGLALPAQVASAQ 30			
DB	164 LIALRPRGATGVAGLANGLALPAQVASAQ 193			
RESULT 2	O8S1V1	PRELIMINARY:	PRT:	424 AA.
ID	O8S1V1:			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Putative dermal glycoprotein.  
 GN P0504E02.8  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 NCBI\_TaxID=39947;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0504E02.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003269; BAB89707.1; -  
 SQ SEQUENCE 424 AA; 44620 MW; 0DADC5CAC8852C34 CRC64;

Query Match 64.8%; Score 94; DB 10; Length 424;  
 Best Local Similarity 67.7%; Pred. No. 0.00014;  
 Matches 21; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Oy 1 LLASLPFGSTGVAGLANSGLALPAQVASAOK 31  
 Db 163 LLAKLPAGAVGAGLARTLALQAVARSOK 193  
 ||| ||| : ||||| : ||| |||| : ||

RESULT 3  
 ID 08S1V3 PRELIMINARY; PRT: 422 AA.  
 AC 08S1V3;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Putative dermal glycoprotein.  
 GN P0504E02.6.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 NCBI\_TaxID=39947;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0504E02.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003269; BAB89705.1; -  
 SQ SEQUENCE 422 AA; 44745 MW; 08CFE26346400A9F CRC64;

Query Match 64.1%; Score 93; DB 10; Length 422;  
 Best Local Similarity 70.0%; Pred. No. 0.00019;  
 Matches 21; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 1 LLASLPFGSTGVAGLANSGLALPAQVASAOK 30  
 Db 165 LLASLPAGAVGAGLGRSLALHAQVAAATQ 194  
 ||||| | : ||||| | : ||| |||| : |

RESULT 4  
 ID 08S1V5 PRELIMINARY; PRT: 434 AA.  
 AC 08S1V5;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Putative dermal glycoprotein.  
 GN P0504E02.3.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.

OX NCBI\_TaxID=39947;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0504E02.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003269; BAB89703.1; -  
 SQ SEQUENCE 434 AA; 45342 MW; 7F2549288A6B28B1 CRC64;

Query Match 60.7%; Score 88; DB 10; Length 434;  
 Best Local Similarity 58.1%; Pred. No. 0.00093;  
 Matches 18; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Oy 1 LLASLPFGSTGVAGLANSGLALPAQVASAOK 31  
 Db 166 LLGSLPGAVGAGLGSAPLSLPQVASAOK 196  
 || ||| | : ||||| : ||| |||| : ||

RESULT 5  
 ID 08S1U4 PRELIMINARY; PRT: 434 AA.  
 AC 08S1U4;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE P0504E02.16 protein.  
 GN P0504E02.16.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 NCBI\_TaxID=39947;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0504E02.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003269; BAB89714.1; -  
 SQ SEQUENCE 434 AA; 46236 MW; 87C36211DD57FE72 CRC64;

Query Match 57.9%; Score 84; DB 10; Length 434;  
 Best Local Similarity 60.7%; Pred. No. 0.0032;  
 Matches 17; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Oy 1 LLASLPFGSTGVAGLANSGLALPAQVASAOK 28  
 Db 167 LLASLPAGATGVAGFSRRLPSLSQALAA 194  
 ||||| | : ||||| | : ||| |||| : |

RESULT 6  
 ID 08S1U6 PRELIMINARY; PRT: 442 AA.  
 AC 08S1U6;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE P0504E02.14 protein.  
 GN P0504E02.14.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 NCBI\_TaxID=39947;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0504E02.";



OS glycine max (Soybean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 NCBI\_TaxID=3847;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. MIYAGISHIROME; TISSUE=DEVELOPING SEEDS;  
 RA Ishizu Y., Sassa H., Hirano H.;  
 RT "Sequence of a cDNA encoding soybean basic 7S globulin isoform."  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB084260; BAB91077.1;  
 KW Signal.  
 FT SIGNAL  
 SO SEQUENCE 433 AA; 47205 MW; FA41B93DBD3EA38 CRC64;  
 Query Match 44.1%; Score 64; DB 10; Length 433;  
 Best Local Similarity 50.0%; Pred. No. 1.6;  
 Matches 14; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
 Oy 1 LLASIPRGSTGVAGLANSGALPAQVAS 28  
 Db 178 LQKGPRIQGVAGIGHAPISLPNOLAS 205  
 RESULT 10  
 09FSZ9 PRELIMINARY: PRT; 369 AA.  
 AC 09FSZ9.  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE Putative extracellular dermal glycoprotein (Fragment).  
 OS Cicier arletium (Chickpea) (Garbanzo).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Ciceraceae; Cicer.  
 NCBI\_TaxID=3827;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. CASTELLANA; TISSUE=ETIOLATED EPICOTYLS;  
 RA Doptico B., Sanchez M.A., Labrador E.;  
 RT "A putative extracellular dermal glycoprotein is expressed in chickpea epicotyls."  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ299060; CAC10209.1;  
 FT NON\_TER  
 SO SEQUENCE 369 AA; 39776 MW; 5DF6FCLDBF3B212B CRC64;  
 Query Match 43.4%; Score 63; DB 10; Length 369;  
 Best Local Similarity 50.0%; Pred. No. 1.9;  
 Matches 13; Conservative 5; Mismatches 8; Indels 0; Gaps 0;  
 Oy 2 LASIPRGSTGVAGLANSGALPAQVA 27  
 Db 109 LNLGPRITGIIIGLARSNLSPQLA 134  
 RESULT 11  
 08SIU9 PRELIMINARY: PRT; 402 AA.  
 AC 08SIU9.  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative dermal glycoprotein.  
 GN P0504E02.10.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrharioideae; Oryzaceae; Oryza.  
 NCBI\_TaxID=39947;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC clone:P0504E02."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP003269; BAB89709.1;  
 SO SEQUENCE 402 AA; 41793 MW; E8C84A214D6F6720 CRC64;  
 Query Match 43.4%; Score 63; DB 10; Length 402;  
 Best Local Similarity 50.0%; Pred. No. 2.1;  
 Matches 14; Conservative 4; Mismatches 10; Indels 0; Gaps 0;  
 Oy 1 LLASIPRGSTGVAGLANSGALPAQVAS 28  
 Db 147 LIRSLPAMAGDAGLGRGVSLPTOLYS 174  
 RESULT 12  
 09ZVS5 PRELIMINARY: PRT; 434 AA.  
 AC 09ZVS5.  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE F15K9.16 protein (Putative extracellular dermal glycoprotein precursor).  
 GN F15K9.16.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Li J., Liu S., Kremenetskaia I., Lueros J., Araujo R., Buehler E., Conway A.B., Dewey K., Feng J., Kim G., Li Y., Shinn P., Sun H., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F15K9 sequence."  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Theologis A.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Theologis A.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Theologis A.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Egu P., Lee J.M., Toriumi M., Yu G., Brooks S., Chao O., Chen H., Karlin-Neumann G., Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene F15K9.16 (GI:3650580)."  
 DL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC005278; AAC72120.1;  
 DR EMBL; AF332411; AAC48774.1;  
 SO SEQUENCE 434 AA; 46148 MW; 17DD684008FDAFC CRC64;  
 Query Match 42.8%; Score 62; DB 10; Length 434;  
 Best Local Similarity 41.4%; Pred. No. 3.1;  
 Matches 12; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T. Matsumoto T. Yamamoto Y.

Job time : 40.3242 secs

Job time : 40.3242 secs

**THIS PAGE BLANK (USPTO)**







Patent No. 6071697  
GENERAL INFORMATION:  
APPLICANT:  
APPLICANT:

TITLE OF INVENTION: NOVEL METHOD FOR TESTING THE  
TITLE OF INVENTION: DIFFERENTIATION STATUS IN PANCREATIC CELLS OF A MAMMAL  
NUMBER OF SEQUENCES: 16  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/778,423A  
FILING DATE: December 31, 1996  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-778-423A-4

Query Match 32.4%; Score 47; DB 3; Length 436;  
Best Local Similarity 34.5%; Pred. No. 27;  
Matches 10; Conservative 8; Mismatches 9; Indels 2; Gaps 1;

QY 4 SLPRGSGV--AGLANSGLAPAOVASA 30  
DB 397 SQPMGTSGTGTGLTSPGVSVQVPGSE 425

## RESULT 9

US-08-115-753-1  
Sequence 1, Application US/08115753  
Patent No. 6017762

## GENERAL INFORMATION:

APPLICANT: JARA, Patrick  
APPLICANT: LOISON, Gerard  
APPLICANT: RAZANAMPARANY, Voahangy  
TITLE OF INVENTION: Cassette for the expression of an  
TITLE OF INVENTION: endothiapsin precursor in Cryphonectria Parasitica  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: P.O. BOX 299  
STREET: King Street Station, Suite 500, 1800 Diagonal Road  
CITY: ALEXANDRIA  
STATE: VIRGINIA  
COUNTRY: USA  
ZIP: 22313-0299

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/115,753  
FILING DATE:

## CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/758872  
FILING DATE: 07-SEP-1991  
APPLICATION NUMBER: FR 9011230  
FILING DATE: 11-SEP-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: SAXE, BERNHARD D  
REGISTRATION NUMBER: 28, 665  
REFERENCE/DOCKET NUMBER: 16781/332  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-9300  
TELEFAX: (703) 683-4109  
TELEX: 899149

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 330 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
IMMEDIATE SOURCE:  
CLONE: endothiapsin  
US-08-115-753-1

Query Match 32.1%; Score 46.5; DB 3; Length 330;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 12; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

QY 9 STGVAGLANSGLALPAOVASA 29  
DB 211 STSIDGLADTGTLLYCPATVSA 234

## RESULT 10

US-09-333-423-4  
Sequence 4, Application US/09333423  
Patent No. 6265636

## GENERAL INFORMATION:

APPLICANT: Randall, Douglas  
APPLICANT: Thelen, Jay  
APPLICANT: Mierzyk, Jan  
APPLICANT: Muszynski, Michael  
APPLICANT: Sewalt, Vincent  
TITLE OF INVENTION: Pyruvate Dehydrogenase Kinase  
TITLE OF INVENTION: Polynucleotides, Polypeptides and Uses Thereof  
FILE REFERENCE: 0818  
CURRENT APPLICATION NUMBER: US/09/333,423  
CURRENT FILING DATE: 1999-06-15  
EARLIER APPLICATION NUMBER: 60/089,998  
EARLIER FILING DATE: 1998-06-19  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 364  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-333-423-4

Query Match 32.1%; Score 46.5; DB 4; Length 364;  
Best Local Similarity 50.0%; Pred. No. 26;  
Matches 12; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

QY 6 PRGSGVAGLANSGLALPAOVASA 29  
DB 175 PKPECVIGLINTRLS-PIQVADA 197

## RESULT 11

US-08-115-753-2  
Sequence 2, Application US/08115753  
Patent No. 6017762

## GENERAL INFORMATION:

APPLICANT: JARA, Patrick  
APPLICANT: LOISON, Gerard  
APPLICANT: RAZANAMPARANY, Voahangy  
TITLE OF INVENTION: Cassette for the expression of an  
TITLE OF INVENTION: endothiapsin precursor in Cryphonectria Parasitica  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: P.O. BOX 299  
STREET: King Street Station, Suite 500, 1800 Diagonal Road  
CITY: ALEXANDRIA  
STATE: VIRGINIA  
COUNTRY: USA  
ZIP: 22313-0299

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/115,753  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/758872  
FILING DATE: 07-SEP-1991  
APPLICATION NUMBER: FR 9011230  
FILING DATE: 11-SEP-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: SAXE, BERNHARD D  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 16781/332  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-9300  
TELEFAX: (703) 683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 419 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
IMMEDIATE SOURCE:  
CLONE: preproendothlepepsin  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 90..419  
US-08-115-753-2

Query Match 32.1%; Score 46.5; DB 3; Length 419;  
Best Local Similarity 50.0%; Pred. No. 31;  
Matches 12; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

OY 9 STGVAGLANS---LALPAQVASA 29  
||:|:|:|:| ||||| ||  
Db 300 STSIDGIADGTTLTLPLATVSA 323

RESULT 12  
US-08-115-753-33  
Sequence 33, Application US/08115753  
Patent No. 6017762  
GENERAL INFORMATION:  
APPLICANT: JARA, Patrick  
APPLICANT: LEGOUX, Richard  
APPLICANT: LOISON, Gerard  
APPLICANT: RAZANAMPARANY, Voahangy  
TITLE OF INVENTION: Cassette for the expression of an  
TITLE OF INVENTION: endothlepepsin precursor in Cryphonectria Parasitica  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: P.O. BOX 299  
STREET: King Street Station, Suite 500, 1800 Diagonal Road  
CITY: ALEXANDRIA  
STATE: VIRGINIA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/115,753  
FILING DATE:  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/758872  
FILING DATE: 07-SEP-1991  
APPLICATION NUMBER: FR 9011230  
FILING DATE: 11-SEP-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: SAXE, BERNHARD D  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 16781/332  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-9300  
TELEFAX: (703) 683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 419 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-115-753-33

Query Match 32.1%; Score 46.5; DB 3; Length 419;  
Best Local Similarity 50.0%; Pred. No. 31;  
Matches 12; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

OY 9 STGVAGLANS---LALPAQVASA 29  
||:|:|:|:| ||||| ||  
Db 300 STSIDGIADGTTLTLPLATVSA 323

RESULT 13  
5310667-9  
Patent No. 5310667  
APPLICANT: EICHHOLTZ, DAVID A.; GASSER, CHARLES S.; KISHORE,  
GANESH M.  
TITLE OF INVENTION: GLYPHOSATE-TOLERANT 5-ENOLPYRUVYL  
-3-PHOSPHOSHIKIMATE SYNTHASES  
NUMBER OF SEQUENCES: 37  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/380,963  
FILING DATE: 17-JUL-1989  
SEQ ID NO: 9;  
LENGTH: 440  
5310667-9

Query Match 31.7%; Score 46; DB 6; Length 440;  
Best Local Similarity 55.6%; Pred. No. 39;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 LLASLRGSGTGVAGLANS 18  
|||:| ||||:| ||:|  
Db 33 LLAALESGSTETGLDLS 50

RESULT 14  
US-08-476-008-64  
Sequence 64, Application US/08476008  
Patent No. 5627061  
GENERAL INFORMATION:  
APPLICANT: Barry, Gerard F.  
APPLICANT: Rishore, Ganesh M.  
APPLICANT: Padgett, Stephen R.  
APPLICANT: Stallings, William C.  
TITLE OF INVENTION: Glyphosate Tolerant  
TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: Dennis R. Hoerner, Jr., Monsanto Co. BBAF  
STREET: 700 Chesterfield Village Parkway  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent'n Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,008  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/306,063  
FILING DATE: 13-SEP-1994  
APPLICATION NUMBER: US 07/749,611  
FILING DATE: 28-AUG-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/576,537  
FILING DATE: 31-AUG-1990  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoerner Jr., Dennis R.  
REGISTRATION NUMBER: 30,914  
REFERENCE/DOCKET NUMBER: 38-21(10660)A  
TELEPHONE: (314)537-6099  
TELEFAX: (314)537-6047  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 442 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-476-008-64

Query Match 31.7%; Score 46; DB 1; Length 442;  
Best Local Similarity 55.6%; Pred. No. 39;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Oy 1 LLASLPRGSGVAGLANS 18  
|||:| ||| :|| :|  
Db 33 LLAALAEGETETGLDLS 50

RESULT 15  
US-08-306-063-64  
Sequence 64, Application US/08306063  
Patent No. 5633435  
GENERAL INFORMATION:  
APPLICANT: Barry, Gerard F.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Padgett, Stephen R.  
APPLICANT: Stallings, William C.  
TITLE OF INVENTION: Glycosate Tolerant  
TITLE OF INVENTION: 5-Bnolpyruvylshikimate-3-phosphate Synthases  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BBAF  
STREET: 700 Chesterfield Village Parkway  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent'n Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/306,063  
FILING DATE: 13-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/749,611

FILING DATE: 28-AUG-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/576,537  
FILING DATE: 31-AUG-1990  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoerner Jr., Dennis R.  
REGISTRATION NUMBER: 30,914  
REFERENCE/DOCKET NUMBER: 38-21(10660)A  
TELEPHONE: (314)537-6099  
TELEFAX: (314)537-6047  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 442 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-306-063-64

Query Match 31.7%; Score 46; DB 1; Length 442;  
Best Local Similarity 55.6%; Pred. No. 39;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Oy 1 LLASLPRGSGVAGLANS 18  
|||:| ||| :|| :|  
Db 33 LLAALAEGETETGLDLS 50

Search completed: May 20, 2003, 16:44:41  
Job time : 15.3077 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:49:49 ; Search time 15.3297 Seconds  
(without alignments)  
200.549 Million cell updates/sec

Title: US-09-869-155-16  
145  
Perfect score: 1  
Sequence: 1 LLASLPGRSTGVAGLANSGLALPAQVASACK 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubppaa/US08\_NEM\_PUB pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/PCU\_NEM\_PUB pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEM\_PUB pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/PCUUS\_PUBCOMB pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_NEM\_PUB pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US10\_NEM\_PUB pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US60\_NEM\_PUB pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	55	37.9	338	9	US-10-219-220-60
2	55	37.9	556	9	US-10-219-220-259
3	48	33.1	319	9	US-09-738-626-5702
4	47.5	32.8	2186	10	US-09-927-668-2
5	46.5	32.1	364	12	US-10-062-254-363
6	46.5	32.1	447	9	US-09-738-626-4148
7	46.5	32.1	1089	9	US-10-174-590-266
8	46.5	32.1	1089	9	US-10-176-758-266
9	46.5	32.1	1089	9	US-10-176-757-266
10	46.5	32.1	1089	9	US-10-176-737-266
11	46.5	32.1	1089	9	US-10-173-706-266
12	46.5	32.1	1089	9	US-10-175-728-266
13	46.5	32.1	1089	9	US-10-175-752-266
14	46.5	32.1	1089	9	US-10-176-482-266
15	46.5	32.1	1089	9	US-10-176-913-266
16	46.5	32.1	1089	9	US-10-180-552-266
17	46.5	32.1	1089	9	US-10-180-557-266
18	46.5	32.1	1089	9	US-10-173-700-266
19	46.5	32.1	1089	9	US-10-174-572-266

20	46.5	32.1	1089	9	US-10-174-579-266	Sequence 266, App
21	46.5	32.1	1089	9	US-10-174-582-266	Sequence 266, App
22	46.5	32.1	1089	9	US-10-174-588-266	Sequence 266, App
23	46.5	32.1	1089	9	US-10-175-739-266	Sequence 266, App
24	46.5	32.1	1089	9	US-10-175-740-266	Sequence 266, App
25	46.5	32.1	1089	9	US-10-175-743-266	Sequence 266, App
26	46.5	32.1	1089	9	US-10-176-488-266	Sequence 266, App
27	46.5	32.1	1089	9	US-10-176-492-266	Sequence 266, App
28	46.5	32.1	1089	9	US-10-176-747-266	Sequence 266, App
29	46.5	32.1	1089	9	US-10-176-750-266	Sequence 266, App
30	46.5	32.1	1089	9	US-10-176-987-266	Sequence 266, App
31	46.5	32.1	1089	9	US-10-176-987-266	Sequence 266, App
32	46.5	32.1	1089	9	US-10-176-991-266	Sequence 266, App
33	46.5	32.1	1089	9	US-10-176-992-266	Sequence 266, App
34	46.5	32.1	1089	9	US-10-176-993-266	Sequence 266, App
35	46.5	32.1	1089	9	US-10-184-638-266	Sequence 266, App
36	46.5	32.1	1089	9	US-10-173-695-266	Sequence 266, App
37	46.5	32.1	1089	9	US-10-173-697-266	Sequence 266, App
38	46.5	32.1	1089	9	US-10-173-705-266	Sequence 266, App
39	46.5	32.1	1089	9	US-10-174-576-266	Sequence 266, App
40	46.5	32.1	1089	9	US-10-174-585-266	Sequence 266, App
41	46.5	32.1	1089	9	US-10-174-586-266	Sequence 266, App
42	46.5	32.1	1089	9	US-10-175-747-266	Sequence 266, App
43	46.5	32.1	1089	9	US-10-176-481-266	Sequence 266, App
44	46.5	32.1	1089	9	US-10-176-485-266	Sequence 266, App
45	46.5	32.1	1089	9	US-10-176-487-266	Sequence 266, App

## ALIGNMENTS

RESULT 1  
US-10-219-220-60  
; Sequence 60, Application US/10219220  
; Publication No. US20030082724A1  
; GENERAL INFORMATION:  
; APPLICANT: Flinn, Barry  
; TITLE OF INVENTION: Compositions affecting programmed cell  
; FILE REFERENCE: 11000.1022c1  
; CURRENT APPLICATION NUMBER: US/10/219,220  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: U.S. NO. US20030082724A1 09/325,932  
; NUMBER OF SEQ ID NOS: 290  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 60  
; LENGTH: 338  
; TYPE: PRT  
; ORGANISM: Pinus radiata  
US-10-219-220-60

Query Match 37.9%; Score 55; DB 9; Length 338;  
Best Local Similarity 44.4%; Pred No. 7;  
Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 LLASLPGRSTGVAGLANSGLALPAQVAS 28  
DB 153 LAVSPAVTGDVGLSSAQVSLPSQLAS 179

RESULT 2  
US-10-219-220-259  
; Sequence 259, Application US/10219220  
; Publication No. US20030082724A1  
; GENERAL INFORMATION:  
; APPLICANT: Flinn, Barry  
; APPLICANT: Lasham, Annette  
; TITLE OF INVENTION: Compositions affecting programmed cell  
; FILE REFERENCE: 11000.1022c1  
; CURRENT APPLICATION NUMBER: US/10/219,220

CURRENT FILING DATE: 2002-08-14  
PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932  
PRIOR FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 290  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 259  
LENGTH: 556  
TYPE: PRT  
ORGANISM: Pinus radiata  
US-10-219-220-259

Query Match 37.9%; Score 55; DB 9; Length 556;  
Best Local Similarity 44.4%; Pred. No. 12;  
Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 LASLP--RGSTGVAGLANSGLALPAQVAS 28  
DB 309 LAVSPAVTDVGLGSSAQVSLPSQLAS 335

RESULT 3  
US-09-738-626-5702  
Sequence 5702, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738, 626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 5702  
LENGTH: 319  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5702

Query Match 33.1%; Score 48; DB 9; Length 319;  
Best Local Similarity 48.3%; Pred. No. 59;  
Matches 14; Conservative 3; Mismatches 10; Indels 2; Gaps 1;

QY 2 LASLP--RGSTGVAGLANSGLALPAQVAS 28  
DB 151 LASLAQVRGVTPAGAFASDVALPSTVA 179

RESULT 4  
US-09-927-668-2  
Sequence 2, Application US/09927668  
Patent No. US20020115144A1  
GENERAL INFORMATION:  
APPLICANT: Kaplan, Jerry  
Perou, Charles  
Moore, Karen  
TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS  
AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME  
NUMBER OF SEQUENCES: 32

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York

COUNTRY: USA  
ZIP: 10036/2711

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/927,668  
FILING DATE: 10-Aug-2001

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/396,540  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-062-999

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2186 amino acids  
TYPE: amino acid

MOLECULE TYPE: protein  
TOPOLOGY: unknown

FRAGMENT TYPE: Internal  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-927-668-2  
Query Match 32.8%; Score 47.5; DB 10; Length 2186;  
Best Local Similarity 50.0%; Pred. No. 5,4e+02;  
Matches 14; Conservative 2; Mismatches 9; Indels 3; Gaps 1;

QY 2 LASLP--RGSTGVAGLANSGLALPAQVASA 29  
DB 621 LASPQRSQSTVASL--GLAFPSQNSGA 645

RESULT 5  
US-10-062-254-363  
Sequence 363, Application US/10062254  
Patent No. US20020138882A1  
GENERAL INFORMATION:  
APPLICANT: Cahoon, Edgar B  
APPLICANT: Cahoon, Rebecca E  
APPLICANT: Falco, Saverio Carl  
APPLICANT: Fang, Yiwen  
APPLICANT: Hantke, Sabine S.  
APPLICANT: Lee, Jian-ming  
APPLICANT: Li, Zhongsen  
APPLICANT: Miao, Guo-hua  
APPLICANT: Morgante, Michele  
APPLICANT: Niu, Xiping  
APPLICANT: Odell, Joan  
APPLICANT: Rafalski, Antoni  
APPLICANT: Sakai, Hajime  
APPLICANT: Zheng, Peizhong  
APPLICANT: Zhu, Qun  
TITLE OF INVENTION: Polynucleotides Encoding Proteins Involved in Plant Metabolism  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/062,254  
CURRENT FILING DATE: 2002-02-01  
PRIOR APPLICATION NUMBER: 09/630,346  
PRIOR FILING DATE: 2000-07-28  
PRIOR APPLICATION NUMBER: 60/146511



```

; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/156006
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/156899
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 60/157287
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 60/169767
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/171054
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: 60/172958
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/171515
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: 60/173535
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 375
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 363
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Zea mays
US-10-062-254-363

Query Match
Best Local Similarity 32.1%; Score 46.5; DB 12; Length 364;
Matches 12; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Db 6 PRGSTVAGLANSGLAPQVASA 29
1: |||||1-: |||||
Db 175 PKPEPGVIGLINTRLS-PIQVQA 197

RESULT 6
US-09-738-626-4148
; Sequence 4148, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKOHO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4148
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4148

Query Match
Best Local Similarity 32.1%; Score 46.5; DB 9; Length 447;
Matches 13; Conservative 3; Mismatches 8; Indels 9; Gaps 1;
```

```

Db 6 GTDVGRLANEVLTAFLALKGAAAHVLTAEK 38

RESULT 7
US-10-174-590-266
; Sequence 266, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 266
; LENGTH: 1089
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-266

Query Match
Best Local Similarity 32.1%; Score 46.5; DB 9; Length 1089;
Matches 13; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

Db 5 LPRGSTVAGLANSGLAL 22
1: |||||1|||1|||
Db 743 LPR---AVAGLANSGLAL 757

RESULT 8
US-10-176-758-266
; Sequence 266, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 266
; LENGTH: 1089
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-266

Query Match
Best Local Similarity 32.1%; Score 46.5; DB 9; Length 1089;
Matches 13; Conservative 0; Mismatches 2; Indels 3; Gaps 1;
```

OY 5 LPRGSGVAGLANSGLAL 22  
||| ||||| |||||  
Db 743 LPR---AVAGLAASGLAL 757

## RESULT 9

US-10-175-737-266  
; Sequence 266, Application US/10175737  
; Publication No. US20030013153A1  
; GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jjian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C50  
CURRENT APPLICATION NUMBER: US/10/175,737  
CURRENT FILING DATE: 2002-06-19  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 266  
LENGTH: 1089  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-175-737-266

Query Match 32.1%; Score 46.5; DB 9; Length 1089;  
Best Local Similarity 72.2%; Pred. No. 3.5e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

OY 5 LPRGSGVAGLANSGLAL 22  
||| ||||| |||||  
Db 743 LPR---AVAGLAASGLAL 757

## RESULT 10

US-10-173-706-266  
; Sequence 266, Application US/10173706  
; Publication No. US20030022293A1  
; GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jjian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C7  
CURRENT APPLICATION NUMBER: US/10/173,706  
CURRENT FILING DATE: 2002-06-17  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 266  
LENGTH: 1089  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-173-706-266

Query Match 32.1%; Score 46.5; DB 9; Length 1089;  
Best Local Similarity 72.2%; Pred. No. 3.5e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

OY 5 LPRGSGVAGLANSGLAL 22  
||| ||||| |||||  
Db 743 LPR---AVAGLAASGLAL 757

## RESULT 11

US-10-175-738-266  
; Sequence 266, Application US/10175738  
; Publication No. US20030022294A1  
; GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jjian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C45  
CURRENT APPLICATION NUMBER: US/10/175,738  
CURRENT FILING DATE: 2002-06-19  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 266  
LENGTH: 1089  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-175-738-266

Query Match 32.1%; Score 46.5; DB 9; Length 1089;  
Best Local Similarity 72.2%; Pred. No. 3.5e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

OY 5 LPRGSGVAGLANSGLAL 22  
||| ||||| |||||  
Db 743 LPR---AVAGLAASGLAL 757

## RESULT 12

US-10-175-752-266  
; Sequence 266, Application US/10175752  
; Publication No. US20030022295A1  
; GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jjian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C60  
CURRENT APPLICATION NUMBER: US/10/175,752  
CURRENT FILING DATE: 2002-06-19  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 266  
LENGTH: 1089  
TYPE: PRT

```
; ORGANISM: Homo Sapien
US-10-175-752-266

Query Match
Best Local Similarity 32.1%; Score 46.5; DB 9; Length 1089;
                          72.2%; Pred. No. 3.5e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

OY 5 LPRGSTGVAGLANSGLAL 22
    ||| |||| ||||
Db 743 LPR---AVAGLAASGLAL 757

RESULT 13
US-10-176-482-266
; Sequence 266, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C70
; CURRENT APPLICATION NUMBER: US/10/176,482
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 266
; LENGTH: 1089
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-482-266

Query Match
Best Local Similarity 32.1%; Score 46.5; DB 9; Length 1089;
                          72.2%; Pred. No. 3.5e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

OY 5 LPRGSTGVAGLANSGLAL 22
    ||| |||| ||||
Db 743 LPR---AVAGLAASGLAL 757

RESULT 14
US-10-176-757-266
; Sequence 266, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C86
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
```

```
; SEQ ID NO 266
; LENGTH: 1089
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-757-266

Query Match
Best Local Similarity 32.1%; Score 46.5; DB 9; Length 1089;
                          72.2%; Pred. No. 3.5e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

OY 5 LPRGSTGVAGLANSGLAL 22
    ||| |||| ||||
Db 743 LPR---AVAGLAASGLAL 757

RESULT 15
US-10-176-913-266
; Sequence 266, Application US/10176913
; Publication No. US20030022298A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C66
; CURRENT APPLICATION NUMBER: US/10/176,913
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 266
; LENGTH: 1089
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-913-266

Query Match
Best Local Similarity 32.1%; Score 46.5; DB 9; Length 1089;
                          72.2%; Pred. No. 3.5e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

OY 5 LPRGSTGVAGLANSGLAL 22
    ||| |||| ||||
Db 743 LPR---AVAGLAASGLAL 757

Search completed: May 20, 2003, 18:00:02
Job time : 16.3297 secs
```

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:04:12 ; Search time 26.1099 Seconds

(without alignments)  
122.483 Million cell updates/sec

Title: US-09-869-155-17

Perfect score: 124

Sequence: 1 GGSPAHYISARFIEVGDRPVSVE 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
A.Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	100.0	24	21	AAV93760
2	124	100.0	24	22	AAE14691
3	124	100.0	24	22	AAU07396
4	112	90.3	381	23	AAU75813
5	105	84.7	381	23	AAU75814
6	104	83.9	381	23	AAU75812
7	101	81.5	196	23	AAU75817
8	97	78.2	370	23	AAU75815
9	97	78.2	370	23	AAU75816
10	75	60.5	185	23	AAU75824

11	75	60.5	185	23	AAU75825
12	75	60.5	185	23	AAU75826
13	75	60.5	277	23	AAU75818
14	65	52.4	102	23	AAU75820
15	65	52.4	102	23	AAU75821
16	61	49.2	287	23	AAU75819
17	51	41.1	170	23	AAU75827
18	49	39.5	218	22	ABB71510
19	48	38.7	410	13	AAU28030
20	48	38.7	862	22	AAU00022
21	47	37.9	366	23	ABP40222
22	45	36.3	89	19	AAW42021
23	45	36.3	89	20	AAU08629
24	45	36.3	89	21	AAU67320
25	45	36.3	540	22	ABB65612
26	45	36.3	1201	20	AAW90345
27	45	36.3	1201	20	ABB58421
28	45	36.3	1348	22	ABB63571
29	44	35.5	122	23	ABP01617
30	44	35.5	208	21	AAU85644
31	44	35.5	329	21	AAU04698
32	44	35.5	485	21	AAU04697
33	44	35.5	496	21	AAU04696
34	43.5	35.1	70	22	AAU28784
35	43.5	35.1	70	23	ABG38068
36	43	34.7	54	22	AAU65550
37	43	34.7	181	23	ABB53437
38	43	34.7	190	22	ABB27458
39	43	34.7	190	22	ABB32608
40	43	34.7	190	22	ABB18104
41	43	34.7	190	22	AAU53437
42	43	34.7	190	22	AAU65815
43	43	34.7	190	22	AAU13677
44	43	34.7	190	22	AAU26076
45	43	34.7	190	22	AAU01427

#### ALIGNMENTS

RESULT 1	AAV93760	standard; peptide; 24 AA.
ID	AAV93760	
AC	AAV93760;	
DT	03-OCT-2000	(first entry)
DE	Amino acid sequence of an endo-beta-1,4-xylanase inhibitor peptide.	
XX	Endo-beta-1,4-xylanase inhibitor; wheat flour; xylanase; bakery product;	
KW	dough; dough preparation.	
OS	Triticum sp.	
XX	WO2000039289-A2.	
PN	06-JUL-2000.	
PD		
XX	17-DEC-1999;	
XX	99WO-1B02071.	
XX	23-DEC-1998;	
PR	99GB-0028599.	
PR	06-APR-1999;	
PR	99GB-0007805.	
PR	15-APR-1999;	
PA	99GB-0008645.	
XX	(DANI-) DANISCO AS.	
XX	Sibbesen O. Sorensen JF;	
PI	WPI; 2000-465744/40.	
DR		
XX	Mutant xylanase protein identified using xylanase inhibitor useful for	
PT	preparing non-sticky dough for bakery products -	

Barley L endoxylan  
Barley L endoxylan  
L endoxylanase inh  
Rye L endoxylanase  
Rye L endoxylanase  
Wheat L endoxylanase  
Wheat L endoxylanase  
Drosophila melanog  
Pichia pastoris pr  
Human activated T  
Staphylococcus epi  
Clone D305\_2 prote  
Human secreted pro  
Drosophila melanog  
Drosophila sp. Cos  
Drosophila melanog  
Drosophila melanog  
Human ORFX protein  
Oxidoreductase ami  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Peptide #2821 enco  
Human peptide enco  
Propionibacterium  
Lactococcus lactis  
Human peptide #109  
Peptide #114 enco  
Protein #103 enco  
Human brain expres  
Human bone marrow  
Peptide #111 enco  
Peptide #113 enco  
Peptide #109 enco

XX Claim 24; Page 112; 112pp; English.  
 PS  
 CC The present sequence is derived from an endo-beta-1,4-xylinase  
 CC inhibitor. The protein is obtained from wheat flour. The specification  
 CC also describes a mutant xylinase protein. The xylinase is useful for  
 CC preparing a foodstuff, preferably a bakery product or a substance  
 CC (e.g. a dough) for making the bakery product. Wild type xylinase or  
 CC mutant xylinase is useful for preparing a dough that is less sticky  
 CC than a dough comprising a fungal xylinase. The xylinase inhibitor is  
 CC useful for screening high degree resistance xylinases for dough  
 CC preparation. The xylinase is also useful for preparing a non-sticky  
 CC dough. A combination of xylinase and the inhibitor is useful for  
 CC calibrating and/or determining the quantity of inhibitor in a wheat  
 CC flour sample.  
 CC  
 XX Sequence 24 AA:  
 SQ  
 Query Match 100.0%; Score 124; DB 21; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-14;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GSPAHYISARFIEGDRVPSVE 24  
 DB 1 GSPAHYISARFIEGDRVPSVE 24  
 RESULT 2  
 ID AAE14691 standard; peptide: 24 AA.  
 AC AAE14691:  
 DT 21-AUG-2002 (first entry)  
 XX  
 DE Wheat flour xylinase inhibitor B chain Lys-C digested fragment #2.  
 XX  
 KW Refrigerated dough; syrruping; arabinoxylan; bakery product; bread;  
 KW pizza base; cake; biscuit; wheat; flour; xylinase inhibitor.  
 XX  
 OS Triticum aestivum.  
 XX  
 PN WO200152657-A1.  
 PD 26-JUL-2001.  
 XX  
 PE 17-JAN-2001; 2001WO-1B00168.  
 XX  
 PR 18-JAN-2000; 2000GB-0001136.  
 PA (DANI-) DANISCO AS.  
 PI Poulsen CH, Sorensen JF;  
 PS WPI: 2001-457446/49.  
 DR  
 PT Production of refrigerated dough with reduced syrruping, useful in  
 PT production of bakery products such as bread, comprises admixing cereal  
 PT flour, water and protein that prevents enzymatic degradation of  
 PT arabinoxylan in the cereal flour -  
 XX  
 XX Disclosure: Page 23; 26pp; English.  
 PS  
 CC The invention relates to a process for producing refrigerated dough  
 CC with reduced 'syrruping' (precipitation of liquid on the dough surface  
 CC because of a reduction in water holding capacity caused by the breakdown  
 CC of arabinoxylan over time). The process comprises admixing cereal flour  
 CC and water with a protein that reduces/prevents enzymatic degradation of  
 CC arabinoxylan in the cereal flour. The preferred protein is a xylinase  
 CC inhibitor. The method is useful to produce refrigerated dough in which  
 CC syrruping is reduced or eliminated. Refrigerated dough is typically  
 CC stored for long periods to enable fresh baked products (e.g. bread,  
 CC pizza bases, cakes, biscuits) to be produced as required. It may also be

CC possible to produce dough or baked products tailored to specific  
 CC requirements using the method by the use of specific proteins/protein  
 CC combinations. The present sequence is wheat flour  
 CC endo-beta-1,4-xylinase inhibitor B chain Lys-C digested fragment.  
 CC  
 XX Sequence 24 AA:  
 SQ  
 Query Match 100.0%; Score 124; DB 22; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-14;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GSPAHYISARFIEGDRVPSVE 24  
 DB 1 GSPAHYISARFIEGDRVPSVE 24  
 RESULT 3  
 ID AAU07396 standard; protein: 24 AA.  
 AC AAU07396:  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Bacillus subtilis xylinase inhibitor #5.  
 XX  
 KW Xylinase; plant cell wall; baking; cereal; starch production; wood;  
 KW wood pulp bleaching; hemicellulose; arabinoxylan; foodstuff; inhibitor.  
 XX  
 OS Bacillus subtilis.  
 XX  
 PN WO20016711-A1.  
 PD 13-SEP-2001.  
 XX  
 PE 08-MAR-2001; 2001WO-1B00426.  
 XX  
 PR 08-MAR-2000; 2000GB-0005585.  
 XX  
 PR 27-JUN-2000; 2000GB-0015751.  
 PA (DANI-) DANISCO AS.  
 PI Sidsesen O, Sorensen JF;  
 PS WPI: 2001-596834/67.  
 DR  
 PT Novel variant xylinase polypeptide or its fragment useful for degrading  
 PT or modifying plant cell wall, comprises amino acid modifications such  
 PT that the polypeptide has altered sensitivity to xylinase inhibitor -  
 XX  
 XX Disclosure: Page 63; 70pp; English.  
 PS  
 CC The invention relates to a variant xylinase polypeptide (I) or its  
 CC fragment having xylinase activity, comprising one or more amino acid  
 CC modifications such that (I) or its fragment has an altered sensitivity to  
 CC a xylinase inhibitor as compared with the parent xylinase enzyme. (I) or  
 CC its coding sequence (II) is useful for degrading or modifying plant cell  
 CC wall or for processing a plant material by contacting the plant cell wall  
 CC or plant material with (I) or (II). (I) is useful for modifying plant  
 CC materials, and in baking, processing cereals, starch production,  
 CC processing wood and enhancing the bleaching of wood pulp. (I)  
 CC is useful for altering the viscosity derived from the presence of  
 CC hemicellulose or arabinoxylan in a solution or system comprising plant  
 CC cell wall material. (I) is useful for preparing a foodstuff such as  
 CC bread, pretzels, tortillas, cakes, cookies, biscuits or crackers.  
 CC The present sequence represents the amino acid sequence of Bacillus  
 CC subtilis xylinase inhibitor #5 as described in the method of the  
 CC invention.  
 XX  
 XX Sequence 24 AA:  
 SQ  
 Query Match 100.0%; Score 124; DB 22; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-14;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGSPAHYSARFIEVGDTRPVSE 24  
 |||||||||||||||||||  
 Db 1 GGSPAHYSARFIEVGDTRPVSE 24

RESULT 4  
 AAU75813  
 ID AAU75813 standard; Protein: 381 AA.

XX AC AAU75813;  
 DT 23-APR-2002 (first entry)  
 XX DE Wheat L endoxylanase inhibitor, TAXI I, variant #2.

XX KW Wheat: TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruiping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

XX OS Triticum aestivum.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 95 /label= unknown  
 FT Misc-difference 98 /label= unknown  
 FT Misc-difference 101 /label= unknown  
 FT Misc-difference 110 /label= unknown  
 FT Misc-difference 110 /label= unknown  
 FT Misc-difference 333 /label= unknown  
 FT Misc-difference 333 /label= unknown  
 XX PN WO200198474-A1.  
 XX PD 27-DEC-2001.  
 XX PF 21-JUN-2001; 2001WO-BE00106.  
 XX PR 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX PA (LEUV-) LEUVEN RES 6 DEV.  
 XX PI Delcours J, Debysse W, Gebruers K, Goesaert H, Flerens K, Robben J;  
 PI Van Campenhout S;  
 DR WPI: 2002-114579/15.  
 XX PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX PS Claim 127; Page 51; 127pp; English.  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,

CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic acid  
 CC acid, a host organism transformed with the nucleic acid of the proteins.  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruiping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a variant of wheat TAXI-I (T. aestivum L endoxylanase  
 CC inhibitor).  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.

XX SQ Sequence 381 AA;  
 XX Query Match 90.3%; Score 112; DB 23; Length 381;  
 XX Best Local Similarity 91.7%; Pred. No. 1.7e-10;  
 XX Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGSPAHYSARFIEVGDTRPVSE 24  
 |||||||||||||||||||  
 Db 201 GGSPAHYSARFIEVGDTRPVSE 224

RESULT 5  
 AAU75814  
 ID AAU75814 standard; Protein: 381 AA.

XX AC AAU75814;  
 XX DT 23-APR-2002 (first entry)  
 XX DE Wheat L endoxylanase inhibitor, TAXI I, microheterogenic variant.

XX KW Wheat: TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruiping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

XX OS Triticum aestivum.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 95 /label= unknown  
 FT Misc-difference 98 /label= unknown  
 FT Misc-difference 101 /label= unknown  
 FT Misc-difference 110 /label= unknown  
 FT Misc-difference 110 /label= unknown  
 FT Misc-difference 145 /label= unknown  
 FT Misc-difference 145 /label= unknown  
 FT Misc-difference 183 /label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys  
 FT Misc-difference 183 /label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys  
 FT Misc-difference 232 /label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys  
 FT Misc-difference 275 /label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys

FT Misc-difference 282  
 FT /label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys  
 FT Misc-difference 333  
 FT /label= Unknown  
 PN W0200198474-AL.  
 XX  
 XX 27-DEC-2001.  
 PD  
 PE 21-JUN-2001; 2001WO-BE00106.  
 XX  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.  
 XX  
 PI Delcours J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 XX  
 DR WPI; 2002-114579/15.  
 XX  
 XX Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 51; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arbinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syripping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arbinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a microheterogenic variant of wheat TAXI-1 ('v. aestivum L  
 CC endoxylanase inhibitor').  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.  
 XX  
 XX  
 XX Sequence 381 AA:  
 SO  
 Query Match 84.7%; Score 105; DB 23; Length 381;  
 Basic Local Similarity 87.5%; Pred. No. 2.7e-09;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 GGSPAHYISARXIEVGDRVPVSE 24

Db 201 GGSPAHYISARXIEVGDRVPVSE 224  
 ||||||||||| ||||||||| 1  
 RESULT 6  
 AAU75812  
 ID AAU75812 standard; Protein: 381 AA.  
 XX  
 AC AAU75812:  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Wheat L endoxylanase inhibitor, TAXI I, variant #1.  
 XX  
 XX Wheat; TAXI-I: L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzymes: dough syripping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 KW  
 XX  
 OS Triticum aestivum.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 95  
 FT /label= Unknown  
 FT Misc-difference 98  
 FT /label= Unknown  
 FT Misc-difference 101  
 FT /label= Unknown  
 FT Misc-difference 110  
 FT /label= Unknown  
 FT Misc-difference 333  
 FT /label= Unknown  
 FT  
 FT W0200198474-AL.  
 XX  
 PD 27-DEC-2001.  
 XX  
 XX 21-JUN-2001; 2001WO-BE00106.  
 PF  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.  
 XX  
 PI Delcours J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 XX  
 DR WPI; 2002-114579/15.  
 XX  
 XX Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 50-51; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arbinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory



CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, wheat xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a variant of wheat TAXI-I (T. aestivum L. endoxylanase  
 CC inhibitor).  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.

CC Sequence 381 AA:

Query Match 83.9%; Score 104; DB 23; Length 381;  
 Best Local Similarity 87.5%; Pred. No. 3.9e-09;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGSPAHYISARFIEVGDTRVPVSE 24  
 |||||  
 Db 201 GGSPAHYISARFIEVGDTRVPVSE 224

RESULT 7

ID AAU75817 standard; Protein: 196 AA.

AC AAU75817;

DT 23-APR-2002 (first entry)

DE Wheat L endoxylanase inhibitor, TDXI I, partial sequence TDXI-I.01.

KW Wheat; TDXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

OS Triticum durum cultivar Mexicali.

PN WO200198474-A1.

PD 27-DEC-2001.

PE 21-JUN-2001; 2001WO-BE00106.

PR 22-JUN-2000; 2000GB-0015296.

PR 25-JAN-2001; 2001GB-0002018.

PR 26-JAN-2001; 2001GB-0002194.

PR 16-MAR-2001; 2001GB-0006564.

PR 21-MAY-2001; 2001GB-0012328.

PA (LEUV-) LEUVEN RES & DEV.

PI Delcourt J, Debysier W, Gebruers K, Goesaert H, Fierens K, Robben J;

DR WPI: 2002-114579/15.

DR N-PSDB; ABRI3674.

PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes

PS Claim 127; Page 57; 127pp; English.

CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulase, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, wheat xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial wheat TDXI-I (T. durum L endoxylanase  
 CC inhibitor).

SO Sequence 196 AA:

Query Match 81.5%; Score 101; DB 23; Length 196;  
 Best Local Similarity 87.0%; Pred. No. 5.7e-09;

Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGSPAHYISARFIEVGDTRVPV 23  
 |||||  
 Db 15 GGSPAHYISARFIEVGDTRVPV 37

RESULT 8

ID AAU75815 standard; Protein: 370 AA.

AC AAU75815;

DT 23-APR-2002 (first entry)

DE Wheat L endoxylanase inhibitor, TAXI I, partial sequence TAXI-I.0.

KW Wheat; TAXI-I; L endoxylanase inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

OS Triticum aestivum cultivar Solissons.

PN WO200198474-A1.

PD 27-DEC-2001.

```

PE 21-JUN-2001; 2001WO-BE00106.
XX
XX 22-JUN-2000; 2000GB-0015296.
PR 25-JAN-2001; 2001GB-0002018.
PR 26-JAN-2001; 2001GB-0002194.
PR 16-MAR-2001; 2001GB-0006564.
PR 21-MAY-2001; 2001GB-0012328.
XX
XX (LEUV-) LEUVEN RES 6 DEV.
PA
XX Delcours J, Delyser W, Gebuwers K, Goesaert H, Flerens K, Robben J;
PI Van Campenhout S;
XX
DR WPI: 2002-114579/15.
DR N-PSDB: ABK13672.
XX
XX Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or
PT beta-glucanolytic enzymes comprises using endoxylanases during
PT screening for inhibition activity or affinity chromatography with
PT immobilised enzymes -
XX
XX Claim 127; Page 57; 127pp; English.
PS
XX
XX The invention relates to separating and/or isolating inhibitors of
CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises
CC screening the inhibition activity by using two or more enzymes during the
CC separation and/or isolation steps that allow to distinguish inhibitors of
CC different specificity or by using an affinity chromatographic step with
CC immobilised enzymes and/or antibodies against inhibitors. Also
CC included are an isolated nucleic acid molecule encoding an inhibitor
CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,
CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,
CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the
CC nucleic acid, an expression system transformed with the nucleic
CC acid, a host organism transformed with the nucleic acid, the inhibitory
CC proteins encoded by the nucleic acids and modulators of the proteins.
CC A recombinant protein, glycoprotein or polypeptide or microorganisms,
CC plant or plant materials transformed with the nucleic acid are useful
CC for the formation of an endoxylanase-inhibitor complex, screening
CC endoxylanases that are totally, less or not inhibited by the inhibitors,
CC reducing syruuping in refrigerated dough compositions, affecting the
CC relative affinity and/or relative hydrolysis specificity and/or relative
CC hydrolysis rate versus water-extractable and/or water-unextractable
CC arabinoxylans of endoxylanases such as by the formation of an
CC endoxylanase/inhibitor complex, improving the maling of cereals such as
CC barley, sorghum and wheat and/or the production of beer, improving the
CC production and/or quality of baked or extruded cereal products such as
CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,
CC biscuits, pasta and noodles, animal feed stuff, improving the production
CC of starch derived syrups, sorbitol, xylitol, xylitol, wheat gluten
CC starch separation and production, improving maize processing,
CC plant disease resistance and nutraceutical and/or pharmaceutical
CC applications, improving paper and pulp technologies. The present
CC sequence represents a partial wheat TAXI-I (T. aestivum L endoxylanase
XX inhibitor).
XX
XX SQ Sequence 370 AA:
XX
XX
XX Query Match 78.2%; Score 97; DB 23; Length 370;
XX Best Local Similarity 83.3%; Pred. No. 66-08;
XX Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0
XX
XX 1 GGSPAHYISARFEVGDTRYPVSVE 24
XX ||||||||| ||||||| |
XX
XX 190 GGSPAHYISARSIWVGDRYPVSE 213
XX
XX
XX RESULT 9
XX ID AAU75816 standard; Protein; 370 AA.
XX AC AAU75816;

```

23-APR-2002 (first entry)

Wheat L endoxylanase inhibitor, TAXI I, partial sequence TAXI-I.02.

Wheat; TAXI-I, L endoxylanase inhibitor; cellulolytic enzyme inhibitor; xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor; immobilised enzyme; enzyme; dough syruiping; cereal product; beer; plant; straight dough; sponge dough; Chorleywood bread; biscuits; pasta; noodle; animal feed; starch separation; maize processing; malting; plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

*Triticum aestivum* cultivar Estica.

WO200198474-A1.

27-DEC-2001.

21-JUN-2001; 2001WO-BE00106.

22-JUN-2000; 2000GB-0015296.

25-JAN-2001; 2001GB-0002018.

26-JAN-2001; 2001GB-0002194.

16-MAR-2001; 2001GB-0006564.

21-MAY-2001; 2001GB-0012328.

(LEUV-) LEUVEN RES & DEV.

Delcours J, Debysers W, Gebruers K, Goesaert H, Fiersens K, Robben J; Van Campenhout S; WPI: 2002-114579/15. N-PSDB; ABR13673.

Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or beta-glucanolytic enzymes comprises using endoxylanases during screening for inhibition activity or affinity chromatography with immobilised enzymes

Claim 127: Page 57; 127pp: English.

The invention relates to separating and/or isolating inhibitors of cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising screening the inhibition activity by using two or more enzymes during the separation and/or isolation steps that allow to distinguish inhibitors of different specificity or by using an affinity chromatographic step with immobilised enzymes and/or antibodies against inhibitors. Also included are an isolated nucleic acid molecule encoding an inhibitor which inhibits cellulase, endoxylanase, beta-glucanase, xylan, alpha-L-arabinofuranosidase and/or other cellulase, xylan, arabinoxylan or beta-glucan degrading enzymes, a vector comprising the nucleic acid, an expression system transformed with the nucleic acid, a host organism transformed with the nucleic acid, the inhibitory proteins encoded by the nucleic acids and modulators of the proteins. A recombinant protein, glycoprotein or polypeptide or microorganisms, plant or plant materials transformed with the nucleic acid are useful for the formation of an endoxylanase-inhibitor complex, screening endoxylanases that are totally, less or not inhibited by the inhibitors, reducing syruiping in refrigerated dough compositions, affecting the relative affinity and/or relative hydrolysis specificity and/or relative hydrolysis rate versus water-extractable and/or water-unextractable arabinoxylans of endoxylanases such as by the formation of an endoxylanase/inhibitor complex, improving the malting of cereals such as barley, sorghum and wheat and/or the production of beer, improving the production and/or quality of baked or extruded cereal products such as straight dough, sponge dough, Chorleywood bread, breakfast cereals, biscuits, pasta and noodles, animal feed stuff, improving the production of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten starch separation and production, improving maize processing, plant disease resistance and nutraceutical and/or pharmaceutical applications, improving paper and pulp technologies. The present sequence represents a partial wheat TAXI-I (T. aestivum L endoxylanase inhibitor).

XX Sequence 370 AA;  
SQ Query Match 78.2%; Score 97; DB 23; Length 370;  
Best Local Similarity 83.3%; Pred. NO. 6e-08;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Oy 1 GSPAHYISARFIEVGTDPVSE 24  
||||||| | | | | | | | |  
Db 190 GSPAHYISARSTVGDTRVPE 213

RESULT 10  
AAU75824  
ID AAU75824 standard; Protein: 185 AA.  
XX AAU75824;  
AC  
XX 23-APR-2002 (first entry)  
DT  
XX  
XX Barley L endoxylanase inhibitor, HVXI I, variant #1.  
DE  
XX  
XX Barley; HVXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
KW immobilised enzyme; enzyme; dough syruing; cereal product; beer; plant;  
KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
KW noodle; animal feed; starch separation; maize processing; malting;  
KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
XX Hordeum vulgare.  
OS

XX Key Location/Qualifiers  
FH Misc-difference 72 /label= Unknown  
FT  
FT Misc-difference 185 /label= Unknown  
FT  
XX  
XX WO200198474-A1.  
XX  
XX 27-DEC-2001.  
XX  
XX 21-JUN-2001; 2001WO-BE00106.  
XX  
XX 22-JUN-2000; 2000GB-0015296.  
XX 25-JAN-2001; 2001GB-0002018.  
XX 26-JAN-2001; 2001GB-0002194.  
XX 16-MAR-2001; 2001GB-0006564.  
XX 21-MAY-2001; 2001GB-0012328.  
XX  
XX (LEUV-) LEUVEN RES & DEV.  
XX  
XX Delcours J, Debysse W, Gebuereers K, Goesaert H, Fierens K, Robben J;  
PI Van Campenhout S;  
XX  
XX WPI: 2002-114579/15.  
XX  
XX Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
PT beta-glucanolytic enzymes comprises using endoxylanases during  
PT screening for inhibition activity or affinity chromatography with  
PT immobilised enzymes  
XX  
XX Claim 127; Page 63; 127pp; English.  
XX  
XX The invention relates to separating and/or isolating inhibitors of  
CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
CC screening the inhibition activity by using two or more enzymes during the  
CC separation and/or isolation steps that allow to distinguish inhibitors of  
CC different specificity or by using an affinity chromatographic step with  
CC immobilised enzymes and/or antibodies against inhibitors. Also  
CC included are an isolated nucleic acid molecule encoding an inhibitor  
CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the

CC nucleic acid, an expression system transformed with the nucleic  
CC acid, a host organism transformed with the nucleic acid, the inhibitory  
CC proteins encoded by the nucleic acids and modulators of the proteins.  
CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
CC plant or plant materials transformed with the nucleic acid are useful  
CC for the formation of an endoxylanase-inhibitor complex, screening  
CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
CC reducing syruing in refrigerated dough compositions, affecting the  
CC relative affinity and/or relative hydrolysis specificity and/or relative  
CC hydrolysis rate versus water-extractable and/or water-unextractable  
CC arabinoxylans of endoxylanases such as by the formation of an  
CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
CC barley, sorghum and wheat and/or the production of beer, improving the  
CC production and/or quality of baked or extruded cereal products such as  
CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
CC biscuits, pasta and noodles, animal feed stuff, improving the production  
CC of starch derived syrups, sorbitol, xylitol, maltitol, wheat gluten  
CC starch separation and production, improving maize processing,  
CC plant disease resistance and nutraceutical and/or pharmaceutical  
CC applications, improving paper and pulp technologies. The present  
CC sequence is a variant of barley HVXI-I (H. vulgare L endoxylanase  
CC inhibitor).  
CC Note: Variant amino acids are highlighted in the specification but  
CC no wild-type sequence is shown for comparison.  
XX  
SQ Sequence 185 AA;  
Query Match 60.5%; Score 75; DB 23; Length 185;  
Best Local Similarity 70.0%; Pred. NO. 0.00015;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
Oy 2 GSPAHYISARFIEVGTDPV 21  
||||| | | | | | | |  
Db 62 GSPAHYISGTIXKVEDTRVP 81

RESULT 11  
AAU75825  
ID AAU75825 standard; Protein: 185 AA.  
XX  
XX AAU75825;  
AC  
XX  
XX 23-APR-2002 (first entry)  
DT  
XX  
XX Barley L endoxylanase inhibitor, HVXI I, variant #2.  
DE  
XX  
XX Barley; HVXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
KW immobilised enzyme; enzyme; dough syruing; cereal product; beer; plant;  
KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
KW noodle; animal feed; starch separation; maize processing; malting;  
KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
XX  
XX Hordeum vulgare.  
OS

XX Key Location/Qualifiers  
FH Misc-difference 72 /label= Unknown  
FT  
FT Misc-difference 185 /label= Unknown  
FT  
XX  
XX WO200198474-A1.  
XX  
XX 27-DEC-2001.  
XX  
XX 21-JUN-2001; 2001WO-BE00106.  
XX  
XX 22-JUN-2000; 2000GB-0015296.  
XX 25-JAN-2001; 2001GB-0002018.  
XX 26-JAN-2001; 2001GB-0002194.  
XX 16-MAR-2001; 2001GB-0006564.  
XX 21-MAY-2001; 2001GB-0012328.  
XX







Search completed: May 20, 2003, 16:14:42  
Job time : 27.109 secs

---

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:11:22 ; Search time 9.75824 Seconds

(without alignments)  
236.439 Million cell updates/sec

Title: US-09-869-155-17

Perfect score: 124

Sequence: 1 GGSPAHYISARFIEVGDTRVPSVE 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries  
PIR\_73:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	48	38.7	136	2 AH3394	lactoylgutathione
2	48	38.7	357	2 F82878	XAA-PRO aminopeptidase
3	48	38.7	682	2 F70421	conserved hypotet
4	47	37.9	161	2 AE0357	conserved hypotet
5	46	37.1	344	2 A95402	probable desaturase
6	46	37.1	444	2 T38760	hypothetical prote
7	46	37.1	455	2 C83494	probable 2-isoprop
8	46	37.1	486	2 A82427	pyruvate kinase II
9	46	37.1	518	2 A13534	probable binding p
10	46	37.1	634	1 S24384	nitrous-oxide redu
11	46	37.1	682	2 T39613	hypothetical prote
12	46	37.1	714	2 T35770	hypothetical prote
13	45.5	36.7	410	2 C84205	hypothetical prote
14	45	36.3	231	2 F70471	hypothetical prote
15	45	36.3	389	2 AD1918	conserved hypotet
16	45	36.3	625	2 B81655	alcohol dehydrogen
17	45	36.3	852	2 T28790	succinate dehydrog
18	45	36.3	955	2 E84022	hypothetical prote
19	45	36.3	1201	2 T08603	hypothetical prote
20	44	35.5	202	2 T46586	kinesin-related pr
21	44	35.5	314	2 T27383	ribosomal protein
22	44	35.5	414	2 E64387	hypothetical prote
23	44	35.5	485	2 C86143	hypothetical prote
24	44	35.5	592	2 T07616	probable beta-fruc
25	44	35.5	636	2 C83222	nitrous-oxide redu
26	44	35.5	918	2 S44769	C25B4.2 protein -
27	43.5	35.1	4930	2 E69679	polyketide synthet
28	43	34.7	148	2 H90257	hypothetical prote
29	43	34.7	165	2 A41072	photosystem I chai

30	43	34.7	181	2 F86641	hypothetical prote
31	43	34.7	247	2 G87472	hypothetical prote
32	43	34.7	282	1 E69906	conserved hypotet
33	43	34.7	459	1 DEPSLP	dihydrolipeamide d
34	43	34.7	464	2 F83365	lipamide dehydrog
35	43	34.7	471	2 S30585	hypothetical prote
36	43	34.7	527	2 E70543	hypothetical prote
37	43	34.7	746	2 T24978	probable L-asparta
38	43	34.7	817	2 T03852	hypothetical prote
39	42.5	34.3	412	1 G70017	protein phosphatas
40	42.5	34.3	781	2 T35029	probable N-carbamy
41	42	33.9	119	2 F42523	hypothetical prote
42	42	33.9	224	2 T10120	A-ORF-A protein -
43	42	33.9	356	2 B84183	F420-dependent NAD
44	42	33.9	376	2 D83099	serine proteinase
45	42	33.9	419	2 AE0937	probable RND efflu
					probable serine hy

#### ALIGNMENTS

RESULT 1  
AH3394  
lactoylgutathione lyase (EC 4.4.1.5) [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 22-Mar-2002  
C:Accession: AH3394  
R:Delevecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov, I.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit  
A:Reference number: A03252; PMID:11756688  
A:Accession: AH3394  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-136 <KUR>  
A:Cross-references: AE008917; PIDN:AAL52323.1; PID:q17963117; GSPDB:GN00190  
C:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BME11142  
A:Map position: 1  
A:Superfamily: hypothetical protein AF2218  
C:Keywords: carbon-sulfur lyase

Query Match 38.7% Score 48: DB 2: Length 136:  
Best Local Similarity 38.1% Pred. No. 2.6:  
Matches 8: Conservative 7: Mismatches 6: Indels 0: Gaps 0:  
OY 4 PAHYISARFIEVGDTRVPSVE 24  
DB 39 PERGVTVFIDVGNKIELLE 59

RESULT 2  
F82878  
XAA-PRO aminopeptidase UN532 [imported] - Ureaplasma urealyticum  
C:Species: Ureaplasma urealyticum  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C:Accession: F82878  
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G  
submitted to GenBank, February 2000  
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views o  
A:Reference number: A82870  
A:Accession: F82878  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-357 <GLA>  
A:Cross-references: GB:AE002152; GB:AF222894; NID:g6899532; PIDN:AAF30945.1; GSP  
A:Experimental source: serovar 3; biovar 1  
C:Genetics:  
A:Gene: pepP; UN532  
A:Genetic code: SGC3  
C:Superfamily: X-Pro aminopeptidase

Query Match 38.7%; Score 48; DB 2; Length 357;  
Best Local Similarity 58.8%; Pred. No. 7.7;  
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 GSPAHYISARFIEVCD 17  
||| | : | | |  
DB 198 GSPHHRHGRNVEDGD 214

## RESULT 3

F70421  
Conserved hypothetical protein aq\_1400 - Aquifex aeolicus

C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 19-Apr-2002

C:Accession: F70421  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: F70421

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-682 <AOF>  
A:Cross-references: GB:AE000738; NID:g2983801; PIDN:AAC07369.1; PID:g2983812; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:  
A:Gene: aq\_1400

C:Superfamily: Acetobacter cellulose synthase operon protein bcsB

Query Match 38.7%; Score 48; DB 2; Length 682;  
Best Local Similarity 31.8%; Pred. No. 16;  
Matches 7; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

OY 3 SPAHYISARFIEVCDTRPSVE 24  
||| | : | | |  
DB 262 TPSSYISEKFVYSDVLPK 283

## RESULT 4

AE0357  
Conserved hypothetical protein YPO2937 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis  
C:Date: 03-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001

C:Accession: AE0357

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B. demo-Terriga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AE0357

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-161 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC92184.1; PID:g1590897; GSPDB:GN00175

C:Genetics:  
A:Gene: YPO2937

Query Match 37.9%; Score 47; DB 2; Length 161;  
Best Local Similarity 45.0%; Pred. No. 4.6;  
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 GSPAHYISARFIEVCDTRV 20  
||| | : | | |  
DB 90 GGDFLEVLTKFTDVIITRV 109

## RESULT 5

A95402  
Probable desaturase [imported] - Sinorhizobium meliloti (strain 1021) megaplasmid pSymA  
C:Species: Sinorhizobium meliloti

C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001

C:Accession: A95402

R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, .; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Y

Proc. Natl. Acad. Sci. U.S.A. 98, 9683-9688, 2001

A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium

A:Reference number: A95262; MUID:21396509; PMID:11481432

A:Accession: A95402

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-344 <KUR>

A:Cross-references: GB:AE006469; PIDN:AAK65779.1; PID:g14524279; GSPDB:GN00165

A:Experimental source: strain 1021, megaplasmid pSymA

R:Gallbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-; pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Le heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K., Ye

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:  
A:Gene: Sma2051

A:Genome: plasmid

Query Match 37.1%; Score 46; DB 2; Length 344;  
Best Local Similarity 47.4%; Pred. No. 16;  
Matches 9; Conservative 4; Mismatches 4; Indels 2; Gaps 1;

OY 5 AHYISARFIEV-GDTRVP 21  
||| | : | | |  
DB 282 AHYLRHXYFEVNGDTRLP 300

## RESULT 6

T38760  
hypothetical protein SPAC3H8.02 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T38760

R:Genies, S.; Churcher, C.M.; Bartell, B.G.; Rajandream, M.A.; Walsh, S.V. submitted to the EMBL Data Library, January 1996

A:Reference number: Z21810

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-444 <GEN>

A:Cross-references: EMBL:269086; PIDN:CAA93159.1; GSPDB:GN00066; SPDB:SPAC3H8.02

A:Experimental source: strain 972h-; cosmid c3H8

C:Genetics:  
A:Gene: SPDB:SPAC3H8.02

A:Map position: 1

Query Match 37.1%; Score 46; DB 2; Length 444;  
Best Local Similarity 52.4%; Pred. No. 21;  
Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

OY 4 PAHYISARFIEVGDTRPSVE 24  
||| | : | | |  
DB 194 PVCYIRARLRHKGDSPEVE 214

## RESULT 7

C83494  
Probable 2-isopropylmalate synthase PAI217 [imported] - Pseudomonas aeruginosa (strain ATCC 27802)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: C83494  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K., L.; Lory, S.; Olson, M.V. Nature 406, 959-964, 2000



C:Accession: T35770  
 R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 Submitted to the EMBL Data Library, July 1998  
 A:Reference number: 221570  
 A:Accession: T35770  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-714 <SSE>  
 A:Cross-references: EMBL:AL031013; PIDN:CAA19775.1; GSPDB:GN00070; SCOPDB:SC8A6.03c  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCOPDB:SC8A6.03c

Query Match 37.1%; Score 46; DB 2; Length 714;  
 Best Local Similarity 42.1%; Pred. No. 35;  
 Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 4 PAHYISARFIEVGDTTRPS 22  
 II::: I I I I I I I I I I  
 Db 327 PARFVAALAEVGRKRIPN 345

RESULT 13  
 C84205  
 hypothetical protein Vng0468c [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: C84205  
 R:Ng, W.V.; Kennedy, S.P.; Mahairs, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabid Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, G.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483; PMID:11016950  
 A:Accession: C84205  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-410 <STO>  
 A:Cross-references: GB:AE004437; NID:g10580073; PIDN:AA019007.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: VNG0468C

Query Match 36.7%; Score 45.5; DB 2; Length 410;  
 Best Local Similarity 42.9%; Pred. No. 23;  
 Matches 9; Conservative 7; Mismatches 4; Indels 1; Gaps 1;

QY 4 PAHYISARFIEVGDTTRPS 24  
 II::: I I I I I I I I I I  
 Db 273 PANFTDNFMAIGDT-VPTVD 292

RESULT 14  
 F70471  
 conserved hypothetical protein aq\_1997 - Aquifex aeolicus  
 C:Species: Aquifex aeolicus  
 C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 17-Mar-2000  
 C:Accession: F70471  
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O V  
 Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A:Reference number: A70300; MUID:98196666; PMID:9537320  
 A:Accession: F70471  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-231 <MO>  
 A:Cross-references: GB:AE000767; NID:g2984235; PIDN:AA007773.1; PID:g2984245; GB:AE00065  
 C:Genetics:  
 A:Experimental source: strain VFS  
 A:Gene: aq\_1997  
 C:Superfamily: carboxymethylenebutenolidase

Query Match 36.3%; Score 45; DB 2; Length 231;  
 Best Local Similarity 42.9%; Pred. No. 15;  
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 GSPAHYISARFIEVGDTTRP 21  
 II::: I I I I I I I I I I  
 Db 120 GGTLMYFAAKPEMVDASLP 140

RESULT 15  
 AD1918  
 alcohol dehydrogenase [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AD1918  
 R:Keneko, T.; Nakamura, Y.; Molk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Iriku Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AD1918  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-389 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA072852.1; PID:g17130241; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: alr0895  
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 36.3%; Score 45; DB 2; Length 389;  
 Best Local Similarity 45.0%; Pred. No. 26;  
 Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 GSPAHYISARFIEVGDTTRP 21  
 I::: I I I I I I I I I I  
 Db 137 GAOAEYIRVPFADVGWVKVP 156

Search completed: May 20, 2003, 16:37:04  
 Job time : 11.7582 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:05:42 ; Search time 6.46154 Seconds  
(without alignments)  
154.055 Million cell updates/sec

Title: US-09-869-155-17  
Perfect score: 124  
Sequence: 1 GGSPAHYISARFIEVGDTRVPSYE 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	37.1	444	YAS2_SCHPO	Q10138 schizosacch
2	46	37.1	634	NS22_PSEAE	Q01710 pseudomonas
3	45	36.3	231	DLHM_AQDAE	Q67802 aquifex aeo
4	44	35.5	202	RS7_NEUCR	Q43105 neurospora
5	44	35.5	414	Y701_MENJA	Q58112 methanococc
6	44	35.5	636	NS21_PSEAE	Q09412 pseudomonas
7	44	35.5	821	YK62_CABEL	P34381 caenorhabdi
8	44	35.5	1385	NAC2_CHURE	Q31686 chlamydomon
9	43	34.7	165	PSAF_SYNY3	P39256 synecocyst
10	43	34.7	181	YBDI_LACIA	Q9c367 lactococcus
11	43	34.7	434	AS15_HUMAN	Q8wxk1 homo sapien
12	43	34.7	459	DLDI_PSEPU	P09063 pseudomonas
13	43	34.7	527	NADB_MYCTU	Q06555 pseudomonas
14	43	34.7	817	NEB2_RAT	Q35274 rattus norv
15	42.5	34.3	412	ALLC_BACCU	Q32149 bacillus su
16	42	33.9	119	YVAA_VACCC	P05010 vaccinia vi
17	42	33.9	419	GLA2_SALTI	Q82229 salmonella
18	42	33.9	662	T9S2_MOUSE	P88021 mus musculu
19	42	33.9	663	T9S2_HUMAN	Q99805 homo sapien
20	42	33.9	1135	RBL2_MOUSE	Q64700 mus musculu
21	42	33.9	1135	RBL2_RAT	Q50501 rattus norv
22	41.5	33.5	436	HEMI_HAUNI	Q3hp12 halobacteri
23	41.5	33.1	115	RUBR_SYNY3	P73068 synecocyst
24	41	33.1	211	THIE_BACDH	Q9ky88 bacillus ha
25	41	33.1	250	LINC_PSEPA	P30159 pseudomonas
26	41	33.1	289	AMIA_SALTY	P33772 salmonella
27	41	33.1	299	PTB_BACSU	P34530 bacillus su
28	41	33.1	390	LPXB_HAEIN	P45011 haemophilus
29	41	33.1	395	KIME_MOUSE	Q91008 mus musculu
30	41	33.1	395	PORA_PYRAB	Q9uy24 pyrococcus
31	41	33.1	431	GLAI_RHIME	Q92q06 rhizobium m
32	41	33.1	448	ACCC_HAEIN	P43873 haemophilus
33	41	33.1	468	GLNA_ECOLI	P06711 escherichia

34	41	33.1	468	1	GLNA_SALTY	P06201 salmonella
35	41	33.1	505	1	GUAA_PYRAE	Q8zt92 pyrobaculum
36	41	33.1	566	1	ROCB_BACSU	P39635 bacillus su
37	41	33.1	1295	1	BXA2_CLOBO	Q45894 clostridium
38	40.5	32.7	331	1	GSPK_AERHY	P31760 aeromonas h
39	40.5	32.7	388	1	GMCR_MOUSE	Q00941 mus musculu
40	40	32.3	125	1	YM07_PARTE	P15608 parametium
41	40	32.3	196	1	RL6_ARCFU	Q28370 archaeoglob
42	40	32.3	208	1	FADD_HUMAN	Q13158 homo sapien
43	40	32.3	251	1	YN31_PYRAE	Q8zy60 pyrobaculum
44	40	32.3	257	1	NUDC_ECO57	Q8x6x7 escherichia
45	40	32.3	257	1	NUDC_ECOLI	P32664 escherichia

## ALIGNMENTS

```

RESULT 1
YAS2_SCHPO          STANDARD:      PRT:      444 AA.
ID  YAS2_SCHPO
AC  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Hypothetical protein C3H8.02 in chromosome I.
GN  SPAC3H8.02.
OS  Schizosaccharomyces pombe (fission yeast).
OC  Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC  Schizosaccharomycetales; Schizosaccharomycetaceae;
OC  Schizosaccharomyces.
OX  NCBI_TaxID=4896;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  SRRATN-972;
RX  MEDLINE=21848401; PubMed=11859360;
RA  Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA  Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA  Brooke K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA  Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA  Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA  Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA  James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA  Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA  Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA  Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA  Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA  Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA  Woodward J., Volkart G., Aert R., Robben J., Gymnopoulos B.,
RA  Weltjens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA  Gabel C., Fuchs M., Fritz C., Holzer E., Moesti D., Hilbert H.,
RA  Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA  Eger P., Zimmermann W., Medler H., Wandt R., Pinnelle B.,
RA  Goffeau A., Cadieu E., Dreano S., Gloux S., Delaere V., Mottier S.,
RA  Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hunt S.M.,
RA  Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA  Daea R.R., Cruzado L., Jimenez J., Sanchez M., Gel Rey F., Benito J.,
RA  Cerrutti L., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA  Shpakovski G.V., Uesely D., Barrell B.G., Nurse P.;
RT  "The genome sequence of Schizosaccharomyces pombe.";
RL  Nature 415:871-880(2002).
-i- SIMILARITY: CONTAINS 1 CRAL-TRIO DOMAIN.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC  EMBL: Z69086; CAA93159.1;
CC  InterPro: IPR001251; CRAL-TRIO.

```



```

ID RS7_NEUCR STANDARD; PRT; 202 AA.
AC 043105;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 40S Ribosomal Protein S7.
GN CRPS-7.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-IVA;
RX MEDLINE=97174111; PubMed=9021131;
RA Vierula P.J.;
RT "Cloning and characterization of a Neurospora crassa ribosomal
protein gene, crps-7."
RL Curr. Genet. 31:139-143(1997).
CC -1 SIMILARITY: BELONGS TO THE S7E FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL: U73847; AAB94301.1;
DR InterPro: IPR000554; Ribosomal_S7E.
DR Pfam: PF01251; Ribosomal_S7e; 1.
DR ProDom: PD006276; Ribosomal_S7E; 1.
DR PROSITE: PS00948; RIBOSOMAL_S7E; 1.
KW Ribosomal protein.
SQ SEQUENCE 202 AA; 22819 MW; 94D87756FEE37798 CRC64;

Query Match 35.5%; Score 44; DB 1; Length 202;
Best Local Similarity 37.9%; Pred. No. 6.8;
Matches 11; Conservative 5; Mismatches 5; Indels 8; Gaps 1;

QY 4 PAHYISARFIEVGDTR-----VPSVE 24
Db 46 PLQFVASREIEVGHGKKAIVIEVPSLQ 74
1 :||| ||| :
1 :||| ||| :

RESULT 5
Y701_METUA STANDARD; PRT; 414 AA.
AC 058112;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0701.
GN MJ0701.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2199;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii."

```

```

RL Science 273:1058-1073(1996).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL: U67516; AAB98698.1;
DR TIGR: MJ0701;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 414 AA; 49101 MW; B53D6FEEB506CAF1 CRC64;

Query Match 35.5%; Score 44; DB 1; Length 414;
Best Local Similarity 53.8%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 GGSPAHYISARF1 13
Db 262 GGSPSHYARSFV 274
||| ||| :
||| ||| :

RESULT 6
NS2L_PSEAE STANDARD; PRT; 636 AA.
AC 09HY12;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrous-oxide reductase precursor (EC 1.7.99.6) (N(2)OR) (N2O
reductase).
GN NOSZ OR PA3392.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01.
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mlitzoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- FUNCTION: NITROUS-OXIDE REDUCTASE IS PART OF A BACTERIAL
RESPIRATORY SYSTEM WHICH IS ACTIVATED UNDER ANAEROBIC
CONDITIONS IN THE PRESENCE OF NITRATE OR NITROUS OXIDE.
CC -1- CATALYTIC ACTIVITY: N(2) + H(2)O + acceptor = nitrous oxide +
reduced acceptor.
CC -1- COFACTOR: COPPER (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER; EACH SUBUNIT CONTAINS 2 DINCLEAR COPPER
CENTERS A AND Z. Z IS THOUGHT TO BE THE SITE OF NITROUS OXIDE
REDUCTION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, TO MITOCHONDRIAL AND
BACTERIAL COX2 SUBUNITS.
CC -1- SIMILARITY: CONTAINS 1 PLASTOCYANIN-LIKE DOMAIN
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----

```







DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ankyrin repeat and SOCS box containing protein 15 (ASB-15) (Fragment).  
 GN ASB15.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kille B.T., Nicola N.A.;  
 RT 'SOCS box proteins'.  
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS AT LEAST 9 ANK REPEATS.  
 CC -1- SIMILARITY: CONTAINS 1 SOCS BOX DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC The European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC DR 3MBL: AF403033; AAL57352.1; -  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR001496; SOCS.  
 DR Pfam: PF000023; ank; 8.  
 DR PRINTS: PR01415; ANKYRIN.  
 DR SMART: SM00248; ANK; 7.  
 DR SMART: SM00253; SOCS; 1.  
 DR PROSITE: PS50088; ANK\_REPEAT; 5.  
 DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE: PS50225; SOCS; 1.  
 DR KW ANK repeat; Repeat.  
 FT NON\_TER 1 1  
 FT REPEAT <1 18 ANK 1.  
 FT REPEAT 22 51 ANK 2.  
 FT REPEAT 55 84 ANK 3.  
 FT REPEAT 88 117 ANK 4.  
 FT REPEAT 121 150 ANK 5.  
 FT REPEAT 153 182 ANK 6.  
 FT REPEAT 195 224 ANK 7.  
 FT REPEAT 225 254 ANK 8.  
 FT REPEAT 262 290 ANK 9.  
 FT DOMAIN 370 425 SOCS BOX.  
 FT SO 434 AA: 48347 MW: E69176A4E01D75C6 CRC64;  
 Query Match 34.7%; Score 43; DB 1; Length 434;  
 Best Local Similarity 50.0%; Pred. No. 23;  
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 OY 8 ISARFIEVDTRVPSV 23  
 DB 252 VNCYFMHVNDRPPSV 267

RX MEDLINE=88329056; PubMed=3046941;  
 RA Burns G., Brown T., Hatter K., Sokatch J.R.;  
 RT "Comparison of the amino acid sequences of the transacylase  
 RT components of branched chain oxoacid dehydrogenase of *Pseudomonas*  
 RT *putida*, and the pyruvate and 2-oxoglutarate dehydrogenases of  
 RT *Escherichia coli*."  
 RL Eur. J. Biochem. 176:165-169(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-PPG2.  
 RX MEDLINE=89137095; PubMed=2917566;  
 RA Burns G., Brown T., Hatter K., Sokatch J.R.;  
 RT "Sequence analysis of the *lpdv* gene for lipamide dehydrogenase of  
 RT branched-chain-oxoacid dehydrogenase of *Pseudomonas putida*."  
 RL Eur. J. Biochem. 179:61-69(1989).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.45 ANGSTROMS).  
 RX MEDLINE=92390345; PubMed=1325638;  
 RA Matevi A., Obmolova G., Sokatch J.R., Betzel C., Hol W.G.J.;  
 RT "The refined crystal structure of *Pseudomonas putida* lipamide  
 RT dehydrogenase complexed with NAD<sup>+</sup> at 2.45-A resolution."  
 RL proteins 13:336-351(1992).  
 CC -1- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX  
 CC CATALYZES THE OVERALL CONVERSION OF ALPHA-KETO ACIDS TO ACYL-COA  
 CC AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMIC COMPONENTS:  
 CC BRANCHED-CHAIN ALPHA-KETO DECARBOXYLASE (E1), LIPAMIDE  
 CC ACYLTRANSFERASE (E2) AND LIPAMIDE DEHYDROGENASE (E3).  
 CC -1- CATALYTIC ACTIVITY: Dihydrolipoamide + NAD(+) = lipoamide + NADH.  
 CC -1- COFACTOR: FAD.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.  
 CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE  
 CC OXIDOREDUCTASES CLASS-1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC The European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC DR EMBL: M57613; AAA65618.1; -  
 DR PIR: S02139; DEPSLP.  
 DR PDB: 1JVL; 31-JAN-94.  
 DR InterPro: IPR001327; FAD\_pyr\_redox.  
 DR InterPro: IPR000205; NAD\_binding.  
 DR InterPro: IPR001100; pyr\_redox.  
 DR InterPro: IPR004099; pyr\_redox\_dim.  
 DR Pfam: PF00070; pyr\_redox; 1.  
 DR Pfam: PF02852; pyr\_redox\_dim; 1.  
 DR PRINTS: PR00368; FADPNR.  
 DR PRINTS: PR00411; PNDRTASE1.  
 DR PRODOM: PD000139; FAD\_pyr\_redox; 1.  
 DR PROSITE: PS00076; PYRIDINE\_REDOX\_1; 1.  
 DR KW Redox-active center; Glycolysis; Oxidoreductase; NAD; Flavoprotein;  
 DR FAD: 3D-structure.  
 FT NP\_BIND 8 38 FAD (ADP PART) (PROBABLE).  
 FT DISULFID 44 49 REDOX-ACTIVE.  
 FT DOMAIN 146 269 NAD-BINDING (BY SIMILARITY).  
 FT NP\_BIND 296 306 FAD (FLAVIN PART) (BY SIMILARITY).  
 FT ACT\_SITE 438 438 BY SIMILARITY.  
 FT STRAND 5 6  
 FT STRAND 9 12  
 FT STRAND 16 28  
 FT HELIX 16 28  
 FT TURN 29 29  
 FT STRAND 32 35  
 FT TURN 40 41  
 FT HELIX 42 47  
 FT TURN 49 49  
 FT HELIX 49 67  
 FT TURN 68 69  
 FT TURN 73 74





GN PUCF.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Bignelli S.C., Bron S.,  
 RA Brouillet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Ertlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,  
 RA Guth S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaser-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Portollik S., Prescott A.M.,  
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipac A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis.";  
 RL Nature 390:249-256(1997).  
 RN [2]  
 RP FUNCTION.  
 RX PubMed=11344136;  
 RA Schultz A.C., Nygaard P., Saxild H.H.;  
 RT "Functional analysis of 14 genes that constitute the purine catabolic  
 RT pathway in Bacillus subtilis and evidence for a novel regulon  
 RT controlled by the PucR transcription activator.";  
 RL J. Bacteriol. 183:3293-3302(2001).  
 CC -I- FUNCTION: INVOLVED IN THE ANAEROBIC UTILIZATION OF ALLANTOIN;  
 CC CONVERTS ALLANTOIN TO (S)-UREIDOGLYCOLATE AND AMMONIA.  
 CC -I- PATHWAY: Degradation of allantoin (purine catabolism); second  
 CC step.  
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M40.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z99120; CAB15243.1; -;  
 DR MEROPS: M40. UNM; -;  
 DR Subtilist: BG13994; pucf.  
 DR InterPro: IPR002933; Peptidase\_M20.  
 DR Pfam: PF01546; Peptidase\_M20; 1.  
 KW Hydrolase; Purine metabolism; Complete proteome.  
 SO SEQUENCE 412 AA; 45519 MW; 852EC0952E4B5802 CRC64;

OY 1 GG---SPAHYISARFIEVG 16  
 II II I III :I:  
 Db 381 GGVSHPKREYTSARQLEIG 399

Search completed: May 20, 2003, 16:35:44  
 Job time : 8.46154 secs

Query Match 34.38; Score 42.5; DB 1; Length 412;  
 Best Local Similarity 52.68; Pred. No. 27;  
 Matches 10; Conservative 2; Mismatches 4; Indels 3; Gaps 1;

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:19:09 ; Search time 29.6703 seconds  
(without alignments)  
166.665 Million cell updates/sec

Title: US-09-869-155-17  
Perfect score: 124  
Sequence: 1 GSPAHYISARFIEVDTRPSVE 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	39.5	218	5 Q9VE39	Q9VE39 drosophila
2	49	39.5	874	10 Q9FLB4	Q9FLB4 arabidopsis
3	48	38.7	136	10 Q8YGL6	Q8YGL6 bruceella me
4	48	38.7	357	16 Q9PPV8	Q9PPV8 ureaplasma
5	48	38.7	682	16 Q67400	Q67400 aquifex aco
6	48	38.7	817	4 Q96SB3	Q96SB3 homo sapien
7	47	37.9	161	16 Q8ZCP1	Q8ZCP1 yersinia pe
8	46.5	37.5	250	16 Q931Y5	Q931Y5 streptomyce
9	46	37.1	344	16 Q92XW9	Q92XW9 rhizobium m
10	46	37.1	455	16 Q914C1	Q914C1 pseudomonas
11	46	37.1	486	16 Q9KLN5	Q9KLN5 vibrio chol
12	46	37.1	509	13 Q8UVU3	Q8UVU3 xenopus lae
13	46	37.1	518	2 Q8VOK9	Q8VOK9 bruceella ab
14	46	37.1	518	16 Q8YDH3	Q8YDH3 schizosacch
15	46	37.1	682	3 Q42947	Q42947 schizosacch
16	46	37.1	714	16 Q87837	Q87837 streptomyce

17	46	37.1	1548	5 Q9NE01	Q9NE01 leishmania
18	45.5	36.7	410	17 Q9HS01	Q9HS01 halobacteri
19	45.5	36.7	984	17 Q97W1	Q97W1 mechanopyru
20	45	36.3	121	16 Q9RLA0	Q9RLA0 streptomyce
21	45	36.3	134	16 Q9KRS3	Q9KRS3 rhizobium l
22	45	36.3	160	12 Q67310	Q67310 influenza vi
23	45	36.3	370	2 Q9X517	Q9X517 uncultured
24	45	36.3	389	16 Q8YF5	Q8YF5 anabaena sp
25	45	36.3	419	10 Q851V0	Q851V0 oryza sativ
26	45	36.3	512	2 Q68181	Q68181 paracoccus
27	45	36.3	522	2 Q68182	Q68182 pseudomonas
28	45	36.3	530	5 Q96671	Q96671 drosophila
29	45	36.3	540	5 Q9VU53	Q9VU53 drosophila
30	45	36.3	625	16 Q9PJE8	Q9PJE8 chlamydia m
31	45	36.3	852	5 Q01978	Q01978 caenorhabdi
32	45	36.3	955	16 Q9K6M5	Q9K6M5 bacillus ha
33	45	36.3	1201	5 Q16644	Q16644 drosophila
34	45	36.3	1201	5 Q9V4K0	Q9V4K0 drosophila
35	45	36.3	1348	5 Q9VAD1	Q9VAD1 drosophila
36	45	36.3	1392	5 Q95TQ7	Q95TQ7 drosophila
37	45	36.3	2864	2 Q8VUE5	Q8VUE5 ervinia chr
38	44.5	35.9	417	5 Q9Y1U6	Q9Y1U6 pristionchu
39	44	35.5	110	2 Q47186	Q47186 escherichia
40	44	35.5	208	2 Q9FAT7	Q9FAT7 paenibacill
41	44	35.5	294	16 Q92L92	Q92L92 rhizobium m
42	44	35.5	298	5 Q9XW86	Q9XW86 caenorhabdi
43	44	35.5	422	10 Q851V3	Q851V3 oryza sativ
44	44	35.5	485	10 Q9LNU3	Q9LNU3 arabidopsis
45	44	35.5	592	10 Q43818	Q43818 pisum sativ

#### ALIGNMENTS

RESULT 1	Q9VE39	PRELIMINARY;	PRT;	218 AA.
ID	Q9VE39			
AC	Q9VE39			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CG7715 protein.			
GN	CG7715.			
OS	Drosophila melanogaster (Fruit fly)			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY;			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.Y.,			
RA	Brannon R.C., Rogers Y.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abell J.F., Adayani A., An H.-G., Andrews-Pfannkoch C., Baldwin D.,			
RA	Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos R., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Fosler A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,			
RA	Hoskins D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,			
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., Moleod M.P., Moberg D.,  
 RA Merklov G., Milshina N.V., Moberg C., Morris J., Moshneff A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazolo M., Piltan G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Slier E., Spadling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003723; AF55590.1; -;  
 DR Flybase: FBgn0038646; CG7715.  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR PRINTS: PR01217; PRICHEXTENSM.  
 SO SEQUENCE 218 AA; 23318 MW; 9F717785433971AE CRC64;

Query Match 39.5%; Score 49; DB 5; Length 218;  
 Best Local Similarity 57.9%; Pred. No. 7.8;  
 Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 3 SPARFIEVGDTRVPSVE 21  
 DB 73 SPARFIEVGDTRVPSVE 91

RESULT 2  
 O9FLB4 PRELIMINARY; PRT; 874 AA.  
 AC O9FLB4;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE NBS/LRR disease resistance protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RX MEDLINE=98344145; PubMed=9679202;  
 RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence  
 RT features of the regions of 1,381,565 bp covered by twenty one  
 RT physically assigned P1 and TAC clones.";  
 RL DNA Res. 5:131-145(1998).  
 DR EMBL: AB010692; BAB0983.1; -;  
 DR InterPro: IPR000767; Disease\_resist.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003591; LRR\_typ.  
 DR InterPro: IPR002182; NB-ARC.  
 DR Pfam: PF00560; LRR; 2.  
 DR Pfam: PF00931; NB-ARC; 1.  
 DR PRINTS: PR00364; DISEASERISIT.  
 DR SMART: SM00369; LRR\_TYP; 2.  
 SO SEQUENCE 874 AA; 99756 MW; 1DD55D0C163B1711 CRC64;

Query Match 39.5%; Score 49; DB 10; Length 874;  
 Best Local Similarity 43.8%; Pred. No. 38;  
 Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 9 SARFIEVGDTRVPSVE 24  
 DB 697 SARFIEVGDTRVPSVE 712

RESULT 3  
 O8YGL6 PRELIMINARY; PRT; 136 AA.  
 AC O8YGL6;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Lactylglutathione lyase (EC 4.4.1.5).  
 GN BHE1142.  
 OS Brucella melitensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29459;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;  
 RX MEDLINE=20020109; PubMed=11756688;  
 RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mufier C., Los T.,  
 RA Ivanova N., Anderson I., Bhattacharya A., Lykdis A., Reznik G.,  
 RA Jablonksi L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,  
 RA Selkov E., Elzer P.H., Hegius S., O'Callaghan D., DeCesson J.-J.,  
 RA Haselkorn R., Kyriades N., Overbeek R.;  
 RT "The genome sequence of the facultative intracellular pathogen  
 RT *Brucella melitensis*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
 DR EMBL: AE009554; AM152323.1; -;  
 DR InterPro: IPR004360; Glyoxalase; 1.  
 DR Pfam: PF00903; Glyoxalase; 1.  
 DR Lyase; Complete proteome.  
 SO SEQUENCE 136 AA; 14580 MW; 7586C1B297240749 CRC64;

Query Match 38.7%; Score 48; DB 16; Length 136;  
 Best Local Similarity 38.1%; Pred. No. 6.5;  
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 4 PAHYISARFIEVGDTRVPSVE 24  
 DB 39 PAHYISARFIEVGDTRVPSVE 59

RESULT 4  
 O9PPV8 PRELIMINARY; PRT; 357 AA.  
 AC O9PPV8;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE XAA-PRO aminopeptidase.  
 GN PEP OR U0532.  
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Molllicutes;  
 OC Mycoplasmataceae; Ureaplasma.  
 OX NCBI\_TaxID=134821;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SEVOVAR 3;  
 RX MEDLINE=20500219; PubMed=11048724;  
 RA Cassell J.I., Letkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
 RT "The complete sequence of the mucosal pathogen *Ureaplasma*  
 RT *urealyticum*.";  
 RL Nature 407:757-762(2000).  
 DR EMBL: AE002152; AAF30945.1; -;  
 DR HSSP: P15034; 1A16.  
 DR InterPro: IPR000994; Peptidase\_M24.  
 DR InterPro: IPR001311; Xaa-Pro-peptids.  
 DR Pfam: PF00557; Peptidase\_M24; 1.  
 DR PROSITE: PS00491; PROLINE\_PEPTIDASE; 1.  
 DR Complete proteome.  
 SO SEQUENCE 357 AA; 39414 MW; EED7D034C791C2C7 CRC64;





RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Kienast H., Kleser H.M., Denapite D., Eichner A., Cullum J.,  
 RA Kinast H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 [4]

RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdeno-Parraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).

CC -!- SIMILARITY: BELONGS TO THE GNTN FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.  
 CC EMBL: AL593842; CAC4213.1; -;  
 DR InterPro: IPR000524; HTH\_GntR.  
 DR Pfam: PF00392; gntR.1.  
 KM DNA-binding: Transcription regulation.

SO SEQUENCE 250 AA; 27022 MW; CA3FE03F64F5CCD6 CRC64;

Query Match 37.5%; Score 46.5; DB 16; Length 250;  
 Best Local Similarity 35.5%; Pred. No. 23;  
 Matches 11; Conservative 2; Mismatches 9; Indels 9; Gaps 1;

OY 1 GSPAH-----YISARIEVGDRVPS 22  
 DB 208 GGVPGHVLLVQRTFYASGRAVETADYVPA 238

RESULT 9

O92XW9 PRELIMINARY; PRT; 344 AA.

AC O92XW9;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Possible desaturase.

GN RAL121 OR SMA2051.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Sinorhizobium.

OX NCBI\_TaxID=382;

RN [1]

RP SEQUENCE FROM N.A.

AC STRAIN-1021;  
 RX MEDLINE=21396509; PubMed=11481432;  
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
 RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gonzy J.,  
 RA Gurjel M., Hong A., Hutzler L., Hyman R.W., Kahn D., Kahn M.L.,  
 RA Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,  
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;  
 RT "Nucleotide sequence and predicted functions of the entire  
 Sinorhizobium meliloti pSyma megaplasmid";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).

DR EMBL: AE007298; AA65779.1; -;  
 DR InterPro: IPR001541; Sterol\_desat.  
 DR Pfam: PF01598; Sterol\_desat; 1.

KW Plasmid; Complete proteome.

SO SEQUENCE 344 AA; 40194 MW; 1C1A72E1139F23E5 CRC64;

Query Match 37.1%; Score 46; DB 16; Length 344;

Best Local Similarity 47.4%; Pred. No. 40;  
 Matches 9; Conservative 4; Mismatches 4; Indels 2; Gaps 1;

OY 5 AHYSARIEV-GDTRVP 21  
 DB 282 AHYDHHKYEYVYGDYLLP 300

RESULT 10

O91AC1 PRELIMINARY; PRT; 455 AA.

AC O91AC1;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Probable 2-isopropylmalate synthase.

GN PA1217.

OC Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI\_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

AC STRAIN-ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig J., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen.";  
 RL Nature 406:959-964(2000).

EMBL: AE004551; AAG04606.1; -;  
 DR InterPro: IPR002034; AIPM/HcIt\_synth.

DR InterPro: IPR000891; HMGL-like.

DR Pfam: PF00682; HMGL-like; 1.

DR PROSITE: PS00816; AIPM\_HOMOCIT\_SYNTH\_2; UNKNOWN\_1;

KW Complete proteome.

SO SEQUENCE 455 AA; 50669 MW; 3B825508D32D8FEF CRC64;

Query Match 37.1%; Score 46; DB 16; Length 455;  
 Best Local Similarity 50.0%; Pred. No. 56;  
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 2 GSPAHYISARIEVGD 17  
 DB 355 GSPFYIAROFHDVGD 370

RESULT 11

O9KLN5 PRELIMINARY; PRT; 486 AA.

AC O9KLN5;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Private kinase II.

GN VCA0708.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OC NCBI\_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

AC STRAIN-EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae.";  
 RL Nature 406:477-483(2000).  
 DR EMBL: AE004400; AAF96607.1; -.  
 DR HSSP: P14178; 1E0T.  
 DR TIGR: VCA0708; -.  
 DR InterPro: IPR001697; Pyruvate\_kinase.  
 DR Pfam: PF00224; PK.C.1.  
 DR Pfam: PF02887; PK.C.1.  
 DR PRINTS: PR01050; PYRUVTKNASF.  
 DR PRODOM: PD001009; Pyruvate\_kinase.1.  
 DR TIGRFAMs: TIGR01064; pyruv\_kin.1.  
 DR Kinase: Complete proteome.  
 KW  
 SQ SEQUENCE 486 AA; 5211 MW; 3D6A239EAD495E0B CRC64;

Query Match 37.1%; Score 46; DB 16; Length 486;  
 Best Local Similarity 58.3%; Pred. No. 60;  
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 13 IEVGDTRPSV 24  
 DB 253 VEIGDARLPSVQ 264

RESULT 12  
 ID 08VU3 PRELIMINARY; PRT; 509 AA.  
 AC 08VU3;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Suci-associated neurotrophic factor target XSNT.  
 GN SNT.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Akai K., Mood K., Daar I.O.;  
 RT "Fibroblast Growth Factor Receptor-Induced Mesoderm Formation in  
 RT Xenopus Embryos is Mediated by Adaptor Protein SNT.";  
 RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF390895; AAL57304.1; -.  
 DR InterPro: IPR002404; Insulin\_ReceptorS1.  
 DR Pfam: PF02174; IRS.1.  
 DR SMART: SM00310; PTB1.1.  
 SQ SEQUENCE 509 AA; 56965 MW; 7FB6CEA4DCD6DF5B CRC64;

Query Match 37.1%; Score 46; DB 13; Length 509;  
 Best Local Similarity 45.5%; Pred. No. 63;  
 Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 2 GSPAHYSARFIEVGDPVSV 23  
 DB 156 GEASSHPSSRHPSVGSTRLPV 177

RESULT 13  
 ID 08VOK9 PRELIMINARY; PRT; 518 AA.  
 AC 08VOK9;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative ABC transporter periplasmic binding protein.  
 GN BATN1953.0RF18.  
 OS Brucella abortus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=235;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=544;  
 RA Bricker B.J.;  
 RT "tn1953, a new element from Brucella abortus.";  
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF454951; AAL59340.1; -.  
 DR InterPro: IPR000914; SBP\_bac-5.  
 DR Pfam: PF00496; SBP\_bac.5.1.  
 DR PROSITE: PS01040; SBP\_BACTERIAL\_5; UNKNOWN\_1.  
 KW  
 SQ SEQUENCE 518 AA; 57014 MW; 9C2PD58CB2C03EE CRC64;

Query Match 37.1%; Score 46; DB 2; Length 518;  
 Best Local Similarity 50.0%; Pred. No. 64;  
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 4 PAHYISARFIEVGDPTR 19  
 DB 503 PAEYINIRFWEIGQAK 518

RESULT 14  
 ID 08YDH3 PRELIMINARY; PRT; 518 AA.  
 AC 08YDH3;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative binding protein YDD5 precursor.  
 GN BMEI10203.  
 OS Brucella melitensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29459;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;  
 RX MEDLINE=20020109; PubMed=11756688;  
 RA Delvecchio V.G., Kapralat V., Redkar R.J., Patra G., Mijer C., Los T.,  
 RA Ivanova N., Anderson I., Bhattacharya A., Lykilds A., Reznik G.,  
 RA Jablonksi L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,  
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
 RA Haselkorn R., Kyrides N., Overbeek R.;  
 RT "The genome sequence of the facultative intracellular pathogen  
 RT Brucella melitensis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
 DR EMBL: AE009658; AAL53444.1; -.  
 DR InterPro: IPR000914; SBP\_bac-5.  
 DR Pfam: PF00496; SBP\_bac.5.1.  
 DR PROSITE: PS01040; SBP\_BACTERIAL\_5; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 518 AA; 57091 MW; BAE42A819BD7238E CRC64;

Query Match 37.1%; Score 46; DB 16; Length 518;  
 Best Local Similarity 50.0%; Pred. No. 64;  
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 4 PAHYISARFIEVGDPTR 19  
 DB 503 PAEYINIRFWEIGQAK 518

RESULT 15  
 ID 042947 PRELIMINARY; PRT; 682 AA.  
 AC 042947;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical 76.7 kDa protein Cl6H5.12C in chromosome II.  
 GN SPBC16H5.12C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OS Schizosaccharomycetes; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Eukaryota;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Wood V., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL022104; CAI1910.2; -;  
DR InterPro: IPR004844; S/T-phosphatase.  
DR Pfam: PF00149; Metallophos; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 682 AA; 76736 MW; 015A5C53C08679C0 CRC64;

Query Match 37.1%; Score 46; DB 3; Length 682;  
Best local Similarity 42.9%; Pred. No. 88;  
Matches 12; Conservative 3; Mismatches 7; Indels 6; Gaps 1;

OY 3 SPAHYIS-----ARFIEVGDTRVPSVE 24  
|| ||: ||: || || ||  
Db 248 SPEHYLOKLSAARAQFMEVYDTVKAIVE 275

Search completed: May 20, 2003, 16:43:18  
Job time : 32.6703 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:35:02 ; Search time 11.0769 Seconds  
(without alignments)  
63.750 Million cell updates/sec

Title: US-09-869-155-17

Perfect score: 124  
Sequence: 1 GSPAHYISARFIEVGDTRPSVE 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	38.7	410	1	US-08-088-633-2
2	48	38.7	410	1	US-08-245-756-2
3	48	38.7	410	1	US-08-441-750-2
4	48	38.7	410	2	US-08-441-751-2
5	48	38.7	410	5	PCT-US92-02521-2
6	47	37.9	366	4	US-09-134-001C-5067
7	45	36.3	89	1	US-08-686-878A-21
8	45	36.3	89	4	US-09-175-928-21
9	45	36.3	1201	4	US-09-098-901-2
10	43	34.7	459	4	US-08-403-545-5
11	43	34.7	459	4	US-08-404-381-5
12	42	33.9	561	2	US-08-756-317-14
13	42	33.9	663	2	US-08-959-004-5
14	41	33.1	10	1	US-08-282-758B-31
15	41	33.1	439	4	US-09-413-814-13
16	41	33.1	508	4	US-08-980-523-9
17	40	32.3	62	4	US-08-894-626-5
18	40	32.3	74	4	US-08-995-159-5
19	40	32.3	74	4	US-08-828-683A-25
20	40	32.3	85	4	US-09-042-785A-28
21	40	32.3	201	4	US-09-064-414-2
22	40	32.3	201	4	US-09-064-414-4
23	40	32.3	208	1	US-08-618-164-3
24	40	32.3	208	4	US-09-382-155-19
25	40	32.3	208	4	US-09-074-044A-19
26	40	32.3	208	4	US-09-064-414-6
27	40	32.3	256	4	US-08-983-502-2

28	40	32.3	256	5	PCT-US95-16542-2	Sequence 2, Appl1
29	40	32.3	256	5	PCT-US96-10521-2	Sequence 2, Appl1
30	40	32.3	760	4	US-09-513-057C-27	Sequence 27, Appl1
31	40	32.3	1802	4	US-09-322-478-18	Sequence 18, Appl1
32	39	31.5	405	4	US-09-134-001C-5259	Sequence 5259, Ap
33	39	31.5	442	4	US-09-357-251-18	Sequence 18, Appl1
34	39	31.5	445	4	US-09-810-671-2	Sequence 2, Appl1
35	38	30.6	162	1	US-07-961-702-2	Sequence 2, Appl1
36	38	30.6	162	1	US-08-472-284-2	Sequence 2, Appl1
37	38	30.6	162	1	US-08-476-678-2	Sequence 2, Appl1
38	38	30.6	162	2	US-08-472-418-2	Sequence 2, Appl1
39	38	30.6	287	4	US-09-371-056-14	Sequence 14, Appl1
40	38	30.6	349	1	US-08-118-270-71	Sequence 71, Appl1
41	38	30.6	349	5	PCT-US93-08528-71	Sequence 8, Appl1
42	38	30.6	389	4	US-09-594-193-8	Sequence 8, Appl1
43	38	30.6	553	2	US-08-586-272-2	Sequence 2, Appl1
44	38	30.6	553	3	US-09-082-969-2	Sequence 2, Appl1
45	38	30.6	612	2	US-08-359-705B-8	Sequence 8, Appl1

## ALIGNMENTS

RESULT 1  
US-08-088-633-2  
; Sequence 2, Application US/08088633  
; Patent No. 5324660  
; GENERAL INFORMATION:  
; APPLICANT: Gleeson, Martin A  
; TITLE OF INVENTION: Genes which influence Pichia proteolytic  
; TITLE OF INVENTION: Genes which influence Pichia proteolytic  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Fitch, Even, Tabin & Flannery  
; STREET: 135 South LaSalle Street, Suite 900  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/088,633  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/678,916  
; FILING DATE: 01-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E  
; REGISTRATION NUMBER: 31192  
; REFERENCE/DOCKET NUMBER: 50848  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619)552-1311  
; TELEFAX: (619)552-0095  
; TELEX: 20 6566 PATLAW CGO  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 410 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-088-633-2  
; Query Match 38.7%; Score 48; DB 1; Length 410;  
; Best Local Similarity. 34.8%; Pred. No. 3.6;  
; Matches 8; Conservative 6; Mismatches 9; Gaps 0;  
; QY 2 GSPAHYISARFIEVGDTRPSVE 24

Db 160 GSMGYSQDVLDIGDLTIPKVD 182

## RESULT 2

US-08-245-756-2  
Sequence 2, Application US/08245756  
Patent No. 5541112  
GENERAL INFORMATION:  
APPLICANT: Gleeson, Martin A  
APPLICANT: Howard, Bradley D  
TITLE OF INVENTION: GENES WHICH INFLUENCE PICHA PROTEOLYTIC  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK  
STREET: 444 So. Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: CA  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/245,756  
FILING DATE: 16-May-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/088,633  
FILING DATE: 06-JULY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/678,916  
FILING DATE: 01-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 9763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-546-4737  
TELEFAX: 619-546-9392  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 410 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-245-756-2

Query Match 38.7%; Score 48; DB 1; Length 410;  
Best Local Similarity 34.8%; Pred. No. 3.6;  
Matches 8; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 2 GSPAHYISARFIEVGDTTPVSE 24

Db 160 GSMGYSQDVLDIGDLTIPKVD 182

US-08-441-750-2

Sequence 2, Application US/08441750  
Patent No. 569166  
GENERAL INFORMATION:  
APPLICANT: Gleeson, Martin A  
APPLICANT: Howard, Bradley D  
TITLE OF INVENTION: GENES WHICH INFLUENCE PICHA PROTEOLYTIC  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street

CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-2926

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/441,750  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/245,756  
FILING DATE: 16-May-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/088,633  
FILING DATE: 06-JULY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/678,916  
FILING DATE: 01-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 9763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 410 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-441-750-2

Query Match 38.7%; Score 48; DB 1; Length 410;  
Best Local Similarity 34.8%; Pred. No. 3.6;  
Matches 8; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 2 GSPAHYISARFIEVGDTTPVSE 24

Db 160 GSMGYSQDVLDIGDLTIPKVD 182

US-08-441-751-2

Sequence 2, Application US/08441751  
Patent No. 5831053  
GENERAL INFORMATION:  
APPLICANT: Gleeson, Martin A  
APPLICANT: Howard, Bradley D  
TITLE OF INVENTION: GENES WHICH INFLUENCE PICHA PROTEOLYTIC  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/441,751  
FILING DATE:  
CLASSIFICATION: 435

```

: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
PCT-US92-02521-2

Query Match          38.7%; Score 48; DB 5; Length 410;
Best Local Similarity 34.8%; Pred. No. 3.6;
Matches 8; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY      2 GSPAHYISARFIEVGDTRVPSVE 24
      || | : | : || | : | :
Db      160 GSMEGYVSQDVLLQIGDILLTPKVD 182

RESULT 6
US-09-134-001C-5067
: Sequence 5067, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: FILE REFERENCE: GFC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: PRIOR FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 5067
: LENGTH: 366
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5067

Query Match          37.9%; Score 47; DB 4; Length 366;
Best Local Similarity 47.6%; Pred. No. 4.7;
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY      1 GSPAHYISARFIEVGDTRVP 21
      || | : | : || || | |
Db      149 GASSESEVRRNGVEVDITVP 169

RESULT 7
US-08-686-878A-21
: Sequence 21, Application US/08686878A
: Patent No. 5708157
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: Lavalley, Edward
: APPLICANT: Ragle, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Evans, Cheryl
: APPLICANT: Spaulding, Vikki
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: TITLE OF INVENTION: ENCODING THEM
: NUMBER OF SEQUENCES: 71
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30

```

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/686,878A  
;; FILING DATE:  
;; CLASSIFICATION: 536  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Brown, Scott A.  
;; REGISTRATION NUMBER: 32,724  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 498-8224  
;; TELEFAX: (617) 876-5851  
;; INFORMATION FOR SEQ ID NO: 21:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 89 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-686-878A-21

Query Match 36.3%; Score 45; DB 1; Length 89;  
Best Local Similarity 72.7%; Pred. No. 1.8;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSPAHYISARF 12  
I:|:|:|:|:|:|  
Db 34 GTPAGYVSARF 44

RESULT 8  
US-09-175-928-21  
; Sequence 21, Application US/09175928A  
; Patent No. 6312921  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavaille, Edward R.  
; APPLICANT: Collins-Racle, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: M. Sha  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: 6006B.A1772A  
; CURRENT APPLICATION NUMBER: US/09/175,928A  
; CURRENT FILING DATE: 1998-10-20  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 21  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (18)  
US-09-175-928-21

Query Match 36.3%; Score 45; DB 4; Length 89;  
Best Local Similarity 72.7%; Pred. No. 1.8;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSPAHYISARF 12  
I:|:|:|:|:|:|  
Db 34 GTPAGYVSARF 44

RESULT 9  
US-09-098-901-2  
; Sequence 2, Application US/09098901B  
; Patent No. 6218144  
; GENERAL INFORMATION:  
; APPLICANT: Scott, Matthew  
; APPLICANT: Sisson, John C.

;; TITLE OF INVENTION: Costal2 Genes and their Uses  
;; FILE REFERENCE: SUN-65P  
;; CURRENT APPLICATION NUMBER: US/09/098,901B  
;; CURRENT FILING DATE: 1998-06-17  
;; EARLIER APPLICATION NUMBER: 60/051,347  
;; EARLIER FILING DATE: 1997-06-30  
;; NUMBER OF SEQ ID NOS: 11  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 2  
;; LENGTH: 1201  
;; TYPE: PRT  
;; ORGANISM: D. Melanogaster  
;; FEATURE:  
;; NAME/KEY: VARIANT  
;; LOCATION: (1)...(1201)  
;; OTHER INFORMATION: xaa = Any Amino Acid  
US-09-098-901-2

Query Match 36.3%; Score 45; DB 4; Length 1201;  
Best Local Similarity 47.6%; Pred. No. 4.5;  
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 GSPAHYISARFIEVCTRPV 21  
I:|:|:|:|:|:|  
Db 117 GSDPMDSIAGLIQVGPHTVP 137

RESULT 10  
US-08-403-545-5  
; Sequence 5, Application US/08403545  
; Patent No. 5656483  
; GENERAL INFORMATION:  
; APPLICANT: Sokatch, John R.  
; APPLICANT: Sykes, Pamela Joy  
; APPLICANT: Madhusudhan, K.T.  
; TITLE OF INVENTION: Genes Encoding Operon and Promoter for  
; TITLE OF INVENTION: Branched Chain Keto Acid Dehydrogenase of Pseudomonas putid  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Carolyn D. Moon  
; ADDRESSEE: Dunlap, Coddling, Perterson and Lee  
; STREET: 9400 N. Broadway, Suite 420  
; CITY: Oklahoma City  
; STATE: Oklahoma  
; COUNTRY: USA  
; ZIP: 73114  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: diskette 5.25 inch, 360 Kb Storage  
; COMPUTER: IBM AT  
; OPERATING SYSTEM: MS-DOS Version 3.3  
; SOFTWARE: Professional Write 2.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/403,545  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 7/603/781  
; APPLICATION NUMBER: 07/172,148  
; FILING DATE: 23-003-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carolyn D. Moon  
; REGISTRATION NUMBER: 33,022  
; REFERENCE/DOCKET NUMBER: 5820.101  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: Attorney, (405) 478-5344  
; TELEFAX: Attorney, (405) 478-5349  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 459 Amino acids  
; TYPE: Amino acid  
; TOPOLOGY: linear



MOLECULE TYPE: Protein  
FEATURE:  
NAME/KEY: Lpd-val, the E3 component  
LOCATION: 4369-5745, N-terminal methionine is present on  
LOCATION: mature protein  
IDENTIFICATION METHOD: Sequence of cyanogen bromide peptides  
Patent No. 5656483  
PUBLICATION INFORMATION:  
AUTHORS: Burns, Gayle  
AUTHORS: Brown, Tracy  
AUTHORS: Hatter, Kenneth  
AUTHORS: Sokatch, John R.  
TITLE: Sequence analysis of the lpdv gene for lipoamide  
Patent No. 5656483  
TITLE: dehydrogenase of Pseudomonas putida  
JOURNAL: European Journal of Biochemistry  
VOLUME: 179  
PAGES: 61-69  
DATE: 1989  
US-08-403-545-5

Query Match 34.7%; Score 43; DB 1; Length 459;  
Best Local Similarity 39.1%; Pred. No. 30;  
Matches 9; Conservative 4; Mismatches 8; Indels 2; Gaps 1;

OY 1 GGSPAHYISARFIEVGDTRVPSV 23  
|||::|||::|||  
Db 13 GGGPGGYVAA--IRAGQLGIPV 33

RESULT 11  
US-08-404-381-5  
Sequence 5, Application US/08404381  
Patent No. 6168945  
GENERAL INFORMATION:  
APPLICANT: Sokatch, John R.  
APPLICANT: Sykes, Pamela Joy  
APPLICANT: Madhusudhan, K. T.  
TITLE OF INVENTION: Genes Encoding Operon and Promoter for  
TITLE OF INVENTION: Branched Chain Keto Acid Dehydrogenase of Pseudomonas putida  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Carolyn D. Moon  
ADDRESS: 9400 N. Broadway, Suite 420  
CITY: Oklahoma City  
STATE: Oklahoma  
COUNTRY: USA  
ZIP: 73114  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 5.25 inch, 360 kb Storage  
COMPUTER: IBM AT  
OPERATING SYSTEM: MS-DOS Version 3.3  
SOFTWARE: Professional Write 2.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/404,381  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/603,781  
FILING DATE:  
APPLICATION NUMBER: 07/172,148  
FILING DATE: 23-003-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Carolyn D. Moon  
REGISTRATION NUMBER: 33,022  
REFERENCE/DOCKET NUMBER: 5820.101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: Attorney, (405) 478-5344  
TELEFAX: Attorney, (405) 478-5349  
TELEX:  
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 459 Amino acids  
TYPE: Amino acid  
TOPOLOGY: Linear  
MOLECULE TYPE: Protein  
FEATURE:  
NAME/KEY: Lpd-val, the E3 component  
LOCATION: 4369-5745, N-terminal methionine is present on  
LOCATION: mature protein  
IDENTIFICATION METHOD: Sequence of cyanogen bromide peptides  
Patent No. 6168945  
PUBLICATION INFORMATION:  
AUTHORS: Burns, Gayle  
AUTHORS: Brown, Tracy  
AUTHORS: Hatter, Kenneth  
AUTHORS: Sokatch, John R.  
TITLE: Sequence analysis of the lpdv gene for lipoamide  
Patent No. 6168945  
TITLE: dehydrogenase of Pseudomonas putida  
JOURNAL: European Journal of Biochemistry  
VOLUME: 179  
PAGES: 61-69  
DATE: 1989  
US-08-404-381-5

Query Match 34.7%; Score 43; DB 4; Length 459;  
Best Local Similarity 39.1%; Pred. No. 30;  
Matches 9; Conservative 4; Mismatches 8; Indels 2; Gaps 1;

OY 1 GGSPAHYISARFIEVGDTRVPSV 23  
|||::|||::|||  
Db 13 GGGPGGYVAA--IRAGQLGIPV 33

RESULT 12  
US-08-756-317-14  
Sequence 14, Application US/08756317  
Patent No. 5849894  
GENERAL INFORMATION:  
APPLICANT: Clemente, Thomas E.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Minsky, Timothy A.  
APPLICANT: Stark, David M.  
TITLE OF INVENTION: Improved Rhodospirillum Rubrum  
TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/756,317  
FILING DATE: 25-NOV-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,693  
FILING DATE: 29-NOV-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Patterson, Melinda L.  
REGISTRATION NUMBER: 33,062  
REFERENCE/DOCKET NUMBER: MOBT:008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (713) 787-1400  
TELEFAX: (713) 787-1440  
TELEX:  
INFORMATION FOR SEQ ID NO: 14:



APPLICANT: Reichenbach, Hans  
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
FILE REFERENCE: PCT/US 99/23535  
CURRENT APPLICATION NUMBER: US/09/413,814  
EARLIER FILING DATE: 1999-10-07  
EARLIER APPLICATION NUMBER: DE 198 46 493.2  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO: 13  
LENGTH: 439  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-413-814-13

Query Match 33.1%; Score 41; DB 4; Length 439;  
Best Local Similarity 41.2%; Pred. No. 62;  
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 8 ISARFIEVGDTRVPSVE 24  
DB 36 VEERFFAGGEDRLPEVE 52

Search completed: May 20, 2003, 16:44:43  
Job time : 13.0769 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4-P5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:49:49 ; Search time 11,8681 Seconds  
(without alignments)  
200,549 Million cell updates/sec

Title: US-09-869-155-17

Perfect score: 124  
Sequence: 1 GSPAHYISARFIEVGDTRVPSVE 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	36.3	89	US-10-114-893-150	Sequence 150, App
2	45	36.3	89	US-10-040-916-21	Sequence 21, App
3	43.5	35.1	70	US-09-864-761-47948	Sequence 47948, A
4	43	34.7	190	US-09-864-761-33402	Sequence 33402, A
5	43	34.7	464	US-09-815-242-11866	Sequence 11866, A
6	41	33.1	390	US-09-815-242-11173	Sequence 11173, A
7	41	33.1	448	US-09-815-242-11160	Sequence 11160, A
8	41	33.1	469	US-09-815-242-10418	Sequence 10418, A
9	41	33.1	469	US-09-815-242-13789	Sequence 13789, A
10	41	33.1	508	US-09-757-415A-1	Sequence 1, Appl
11	41	33.1	538	US-09-758-269-8	Sequence 8, Appl
12	41	33.1	538	US-09-758-269-18	Sequence 18, Appl
13	41	33.1	538	US-09-758-269-33	Sequence 33, Appl
14	41	33.1	915	US-10-029-180-74	Sequence 74, Appl
15	40.5	32.7	607	US-09-344-882-18	Sequence 18, Appl
16	40	32.3	40	US-09-764-891-5300	Sequence 5300, Ap
17	40	32.3	62	US-10-035-408-5	Sequence 5, Appl
18	40	32.3	74	US-10-112-793-25	Sequence 25, Appl
19	40	32.3	176	US-10-168-066-6	Sequence 6, Appl

20	40	32.3	256	10	US-09-933-814-2	Sequence 2, Appl
21	40	32.3	256	10	US-09-824-134-2	Sequence 2, Appl
22	40	32.3	383	9	US-09-738-626-4233	Sequence 4233, Ap
23	40	32.3	394	9	US-09-738-626-4273	Sequence 4273, Ap
24	40	32.3	760	10	US-09-746-801A-27	Sequence 27, Appl
25	40	32.3	842	10	US-09-798-831-8	Sequence 8, Appl
26	40	32.3	1802	10	US-09-965-553-18	Sequence 18, Appl
27	39.5	31.9	765	9	US-10-078-770-84	Sequence 94, Appl
28	39.5	31.9	818	9	US-10-078-770-94	Sequence 94, Appl
29	39	31.5	159	9	US-09-847-208-35	Sequence 35, Appl
30	39	31.5	159	9	US-09-847-208-41	Sequence 41, Appl
31	39	31.5	419	9	US-09-738-626-6634	Sequence 6634, Ap
32	39	31.5	445	12	US-10-109-854-2	Sequence 2, Appl
33	38.5	31.0	86	9	US-09-764-891-3367	Sequence 3657, Ap
34	38.5	31.0	106	9	US-09-764-891-3390	Sequence 3990, Ap
35	38.5	31.0	411	9	US-10-125-540-351	Sequence 351, App
36	38.5	31.0	411	10	US-09-764-870-351	Sequence 351, App
37	38	30.6	66	9	US-09-764-891-5186	Sequence 5186, Ap
38	38	30.6	68	9	US-09-764-891-5313	Sequence 5313, Ap
39	38	30.6	91	10	US-09-864-761-36795	Sequence 36795, A
40	38	30.6	178	9	US-09-798-889-76	Sequence 96, Appl
41	38	30.6	238	9	US-10-078-770-96	Sequence 96, Appl
42	38	30.6	265	10	US-09-815-242-5157	Sequence 5157, Ap
43	38	30.6	287	12	US-10-081-301-14	Sequence 14, Appl
44	38	30.6	298	10	US-09-815-242-5142	Sequence 5142, Ap
45	38	30.6	305	9	US-10-078-770-90	Sequence 90, Appl

#### ALIGNMENTS

RESULT 1  
US-10-114-893-150  
; Sequence 150, Application US/10114893  
; Publication No. US20020193567A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallee, Edward R.  
; APPLICANT: Collins-Racle, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Bowman, Michael R.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Carlin-Duckett, McKeough  
; APPLICANT: Kelleher, Kerry S.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: GI 6000-10A  
; CURRENT APPLICATION NUMBER: US/10/114,893.  
; CURRENT FILING DATE: 2002-04-02  
; EARLIER APPLICATION NUMBER: 09/413,232  
; EARLIER FILING DATE: 1999-10-06  
; NUMBER OF SEQ ID NOS: 321  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 150  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (18)  
US-10-114-893-150

Query Match 36.3%; Score 45; DB 9; Length 89;  
Best local Similarity 72.7%; Pred. No. 7.1;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSPAHYISARF 12  
I: I I I I I I I I  
DB 34 GTPAGYVSARF 44

RESULT 2  
US-10-040-916-21  
Sequence 21, Application US/10040916  
Patent No. US20020146769A1  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
McCoy, John  
Lavallee, Edward  
Racie, Lisa  
Merberg, David  
Treaacy, Maurice  
Evans, Cheryl  
Spaulding, Vikki  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
ENCODING THEM  
NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/040,916  
FILING DATE: 07-Jan-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/887,029  
FILING DATE: 07-FEB-1997  
APPLICATION NUMBER: 08/686,878  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 89 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 21:  
US-10-040-916-21  
Query Match 36.3%; Score 45; DB 12; Length 89;  
Best Local Similarity 72.7%; Pred. No. 7.1;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GSPAHYSARF 12  
DB 34 GTPAGTYSARF 44  
RESULT 3  
US-09-864-47948  
Sequence 47948, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aecmca-x-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US-60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 47948  
LENGTH: 70  
TYPE: prt  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC006042.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1  
OTHER INFORMATION: EST\_HUMAN HIT: R45715.1, EVALU 9.00e-19  
US-09-864-761-47948  
Query Match 35.1%; Score 43.5; DB 10; Length 70;  
Best Local Similarity 47.6%; Pred. No. 9.4;  
Matches 10; Conservative 4; Mismatches 6; Indels 1; Gaps 1;  
QY 4 PAHYISARFIEVGDTRPSVE 24  
DB 13 PAHYRSSSTRSI-DTQPSVO 32  
RESULT 4  
US-09-864-761-33402  
Sequence 33402, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

```
FILE REFERENCE: Aecmica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 33402
LENGTH: 190
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006333.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 2.5
OTHER INFORMATION: EST_HUMAN HIT: BE466332.1, EVALUE 8.00e-36
OTHER INFORMATION: SWISSPROT HIT: Q02357, EVALUE 6.00e-08
US-09-864-761-33402

Query Match          34.7% Score 43: DB 10: Length 190:
Best Local Similarity 50.0% Pred. No. 34:
Matches 8: Conservative 3: Mismatches 5: Indels 0: Gaps 0:
```

```
QY      8 ISARFLEVDGTRVPSV 23
      : : | | | | |
DB     116 VNCYFMHVDTRPSV 131

RESULT 5
US-09-815-242-11866
Sequence 11866, Application US/09815242
```

```
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes In
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11866
LENGTH: 464
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11866

Query Match          34.7% Score 43: DB 10: Length 464:
Best Local Similarity 39.1% Pred. No. 93:
Matches 9: Conservative 4: Mismatches 8: Indels 2: Gaps 1:

QY      1 GGSFPAHYISARFLEVDGTRVPSV 23
      : | | | | | | | | | | | | |
DB     13 GGGPGGYVAA--IRAGQLGIPV 33

RESULT 6
US-09-815-242-11173
Sequence 11173, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes In
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
```

```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11173
; LENGTH: 390
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
US-09-815-242-11173
```

```

Query Match          33.1%; Score 41; DB 10; Length 390;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 5 AHHSARFIEVGDT 20
    |||:||||:|:
DB 30 AHYNARFIEIGAPRM 45
```

```

RESULT 7
US-09-815-242-11160
; Sequence 11160, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11160
; LENGTH: 448
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
US-09-815-242-11160
```

```

Query Match          33.1%; Score 41; DB 10; Length 448;
Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 10; Conservative 1; Mismatches 5; Indels 2; Gaps 1;
```

```
QY 4 PAHYIS-ARFIEVGDT 19
    |||:|:|:|:|:|:
DB 378 PPHYDSMIKLTLYGDT 395
```

```

RESULT 8
US-09-815-242-10418
```

```

; Sequence 10418, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10418
; LENGTH: 469
; TYPE: PRF
; ORGANISM: Escherichia coli
US-09-815-242-10418
```

```

Query Match          33.1%; Score 41; DB 10; Length 469;
Best Local Similarity 53.8%; Pred. No. 1.9e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 4 PAHYISARFIEVG 16
    |||:|:|:|:|:
DB 35 PAHVNAEFEEEG 47
```

```

RESULT 9
US-09-815-242-13789
; Sequence 13789, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
```



RESULT 11  
 US-09-758-269-8  
 Sequence 8, Application US/09758269  
 Patent No. US20020104120A1  
 GENERAL INFORMATION:  
 APPLICANT: IUCHI, SATOSHI  
 APPLICANT: KOBAYASHI, MASATOMO  
 APPLICANT: SHINOZAKI, KAZUO  
 TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING  
 TITLE OF INVENTION: CLEAVAGE ENZYME GENE  
 FILE REFERENCE: 3914-3  
 CURRENT APPLICATION NUMBER: US/09/758, 269  
 CURRENT FILING DATE: 2001-01-12  
 PRIOR APPLICATION NUMBER: JP 2001-003476  
 PRIOR FILING DATE: 2001-01-11  
 PRIOR APPLICATION NUMBER: JP 2000-010056  
 PRIOR FILING DATE: 2000-01-13  
 NUMBER OF SEQ ID NOS: 33  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 8  
 LENGTH: 538

```

Query Match      33.1%;  Score 41;  Db 10;  Length 538

US-09-758-269-33
; RESULT 13
; Sequence 33, Application US/09/58269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHIMIZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVEAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758, 269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 538
; TYPE: prnt
; ORGANISM: Arabidopsis thaliana
US-09-758-269-33

```

Best Local Similarity 38.9%; Pred. No. 2.3e+02;  
Matches 7; Conservative 2; Mismatches 9; Indels 0; Gaps 0;  
QY 4 PAHYISARFIEVGDTRVP 21  
DB 44 PLHYLSGNFAPIRDERPP 61

## RESULT 14

US-10-029-180-74  
Sequence 74, Application US/10029180  
Publication No. US20020182708A1  
GENERAL INFORMATION:  
APPLICANT: Call, Brian M.  
APPLICANT: Holtzman, Doug  
APPLICANT: Madden, Kevin T.  
APPLICANT: Milna, G. Todd  
APPLICANT: Sherman, Amir  
APPLICANT: Silva, Jeffrey C.  
APPLICANT: Trueheart, Josh  
APPLICANT: Zhang, Lixin  
TITLE OF INVENTION: No. US20020182708A1el Regulators of Fungal Gene Expression  
FILE REFERENCE: MIC-004  
CURRENT APPLICATION NUMBER: US/10/029,180  
CURRENT FILING DATE: 2001-12-22  
PRIOR APPLICATION NUMBER: US 60/257,431  
PRIOR FILING DATE: 2000-12-22  
NUMBER OF SEQ ID NOS: 138  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 74  
LENGTH: 915  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: fungal gene  
US-10-029-180-74

Query Match 33.1%; Score 41; DB 9; Length 915;  
Best Local Similarity 39.1%; Pred. No. 4.1e+02;  
Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 GGSFAHYISARFIEVGDTRVP 23  
DB 36 GSAPHEHMSFDIKVDVIAVPNV 58

## RESULT 15

US-09-344-882-18  
Sequence 18, Application US/09344882  
Patent No. US20020162137A1  
GENERAL INFORMATION:  
APPLICANT: Nikolau, Basil J  
APPLICANT: Wurtel, Eve S  
APPLICANT: Oliver, David J  
APPLICANT: Behal, Robert  
APPLICANT: Schnable, Patrick S  
APPLICANT: Ke, Jinsan  
APPLICANT: Johnson, Jerry L  
APPLICANT: Allred, Carolyn C  
APPLICANT: Faliand, Beth  
APPLICANT: Lutziger, Isabelle  
APPLICANT: Wen, Tsui-Jung  
TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and  
TITLE OF INVENTION: Acetyl CoA Levels in Plants  
FILE REFERENCE: 201573  
CURRENT APPLICATION NUMBER: US/09/344,882  
CURRENT FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: US 60/090,717  
PRIOR FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.2  
SEQ ID NO 18  
LENGTH: 607

TYPE: PRT  
ORGANISM: Arabidopsis Thaliana  
US-09-344-882-18

Query Match 32.7%; Score 40.5; DB 9; Length 607;  
Best Local Similarity 32.4%; Pred. No. 3.1e+02;  
Matches 11; Conservative 4; Mismatches 6; Indels 13; Gaps 1;

QY 3 SPAH-----YISARFIEVGDTRVP 23  
DB 33 SPLHSTTVSPCDATLGRYLARLRLVEIGVTDVPSV 66

Search completed: May 20, 2003, 18:00:03  
Job time : 12.8681 secs

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:04:12 ; Search time 14.1429 Seconds  
(without alignments)  
122.483 Million cell updates/sec

Title: US-09-869-155-18

Perfect score: 65  
Sequence: 1 VNVGVLAACAPSK 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq.101002:\*

- 1: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SIDS2/gcgcdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	100.0	13	21	AAV93761
2	65	100.0	13	22	AAE14692
3	65	100.0	13	22	AAU07397
4	65	100.0	370	23	AAU75815
5	65	100.0	370	23	AAU75816
6	65	100.0	381	23	AAU75812
7	65	100.0	381	23	AAU75813
8	65	100.0	381	23	AAU75814
9	50	76.9	287	23	AAU75819
10	43	66.2	277	23	AAU75818

11	41	63.1	25	20	AAW81605	Synthetic rabies N
12	39	60.0	25	9	AAE08689	Sequence correxp.
13	39	60.0	25	10	AAE91478	Peptide N-V12b. R
14	39	60.0	25	20	AAW81603	Synthetic peptide
15	39	60.0	341	21	AAE42355	Arabidopsis thalia
16	39	60.0	342	21	AAE42354	Arabidopsis thalia
17	39	60.0	450	8	AAE70317	Protein N encoded
18	38	58.5	26	20	AAW81612	Synthetic peptide
19	38	58.5	35	22	ABE31621	Peptide #4272 enco
20	38	58.5	35	22	ABE36838	Peptide #4344 enco
21	38	58.5	35	22	ABE22161	Protein #4160 enco
22	38	58.5	35	22	AAE57582	Human brain expres
23	38	58.5	35	22	AAE69586	Human bone marrow
24	38	58.5	35	22	AAE17817	Peptide #4251 enco
25	38	58.5	35	22	AAE03024	Peptide #4361 enco
26	38	58.5	35	22	AAE05464	Peptide #4146 enco
27	38	58.5	35	23	ABE39615	Human peptide enco
28	38	58.5	63	22	AAE6672	Novel human connec
29	38	58.5	197	21	AAE41482	Human OREX ORF1246
30	38	58.5	197	23	ABE32724	Human ORE1697 prot
31	37	56.9	116	22	ABE11090	Human hydrophobic
32	37	56.9	529	22	AAU4300	Propionibacterium
33	36	55.4	54	22	ABE24978	Novel human diagno
34	36	55.4	89	20	AAE35915	Extended human sec
35	36	55.4	159	22	ABE15894	Novel human diagno
36	36	55.4	333	21	ABE58749	Arabidopsis thalia
37	36	55.4	361	22	ABE04048	Novel human diagno
38	36	55.4	369	21	AAE52007	Arabidopsis thalia
39	36	55.4	377	21	AAE53880	A rabbit CD14 prot
40	36	55.4	390	20	AAE37383	Amino acid sequenc
41	36	55.4	445	21	AAE17005	Arabidopsis thalia
42	36	55.4	445	21	AAE52006	Arabidopsis thalia
43	36	55.4	459	20	AAE38604	Neisseria meningit
44	36	55.4	478	22	AAU30648	Novel human secret
45	36	55.4	486	22	ABE14794	Novel human diagno

#### ALIGNMENTS

RESULT 1  
ID AAV93761 standard; peptide: 13 AA.  
XX AAV93761:  
AC  
XX  
XX  
DT 03-OCT-2000 (first entry)  
XX  
DE Amino acid sequence of an endo-beta-1,4-xylanase inhibitor peptide.  
XX  
KW Endo-beta-1,4-xylanase inhibitor; wheat flour; xylanase; bakery product;  
XX  
XX dough; dough preparation.  
OS Triticum sp.  
XX  
XX WO200039289-A2.  
XX  
XX 06-JUL-2000.  
XX  
XX  
XX 17-DEC-1999; 99WO-IB02071.  
XX  
XX 23-DEC-1998; 98GB-0028599.  
XX  
XX 06-APR-1999; 99GB-0007805.  
XX  
XX 15-APR-1999; 99GB-0008645.  
XX  
XX (DANT-) DANISCO AS.  
XX  
XX Sibbesen O, Sorensen JF;  
XX  
XX WPL; 2000-465744/40..  
XX  
XX Mutant xylanase protein identified using xylanase inhibitor useful for  
XX preparing non-sticky dough for bakery products .  
PT

XX Claim 24; Page 112; 112pp; English.  
PS  
XX  
CC The present sequence is derived from an endo-beta-1,4-xylanase  
CC inhibitor. The protein is obtained from wheat flour. The specification  
CC also describes a mutant xylanase protein. The xylanase is useful for  
CC preparing a foodstuff, preferably a bakery product or a substance  
CC (e.g. a dough) for making the bakery product. Wild type xylanase or  
CC mutant xylanase is useful for preparing a dough that is less sticky  
CC than a dough comprising a fungal xylanase. The xylanase inhibitor is  
CC useful for screening high degree resistance xylanases for dough  
CC preparation. The xylanase is also useful for preparing a non-sticky  
CC dough. A combination of xylanase and the inhibitor is useful for  
CC calibrating and/or determining the quantity of inhibitor in a wheat  
CC flour sample.  
XX  
SQ Sequence 13 AA;  
Query Match 100.0%; Score 65; DB 21; Length 13;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VNWGVLAACAPSK 13  
Db 1 VNWGVLAACAPSK 13  
RESULT 2  
AE14692 standard; peptide: 13 AA.  
ID AE14692 standard; peptide: 13 AA.  
XX  
AC AE14692:  
XX  
DT 21-AUG-2002 (first entry)  
XX  
DE wheat flour xylanase inhibitor B chain Lys-C digested fragment #3.  
XX  
KM Refrigerated dough; syruping; arabinoxylan; bakery product; bread;  
KM pizza base; cake; biscuit; wheat; flour; xylanase inhibitor.  
XX  
OS Triticum aestivum.  
XX  
PN WO200152657-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 17-JAN-2001; 2001WO-1B00168.  
XX  
PR 18-JAN-2000; 2000GB-0001136.  
XX  
PA (DANI-) DANISCO AS.  
XX  
PI Sorensen CH, Sorensen JF;  
XX  
DR WPI; 2001-457446/49.  
XX  
PT Production of refrigerated dough with reduced syruping, useful in  
PT production of bakery products such as bread, comprises admixing cereal  
PT flour, water and protein that prevents enzymatic degradation of  
PT arabinoxylan in the cereal flour -  
XX  
PS Disclosure; Page 23; 26pp; English.  
XX  
CC The invention relates to a process for producing refrigerated dough  
CC with reduced 'syruping' (precipitation of liquid on the dough surface  
CC because of a reduction in water holding capacity caused by the breakdown  
CC of arabinoxylan over time). The process comprises admixing cereal flour  
CC and water with a protein that reduces/prevents enzymatic degradation of  
CC arabinoxylan in the cereal flour. The preferred protein is a xylanase  
CC inhibitor. The method is useful to produce refrigerated dough in which  
CC syruping is reduced or eliminated. Refrigerated dough is typically  
CC stored for long periods to enable fresh baked products (e.g. bread,  
CC pizza bases, cakes, biscuits) to be produced as required. It may also be

CC possible to produce dough or baked products tailored to specific  
CC requirements using the method by the use of specific proteins/protein  
CC combinations. The present sequence is wheat flour  
CC endo-beta-1,4-xylanase inhibitor B chain Lys-C digested fragment.  
XX  
SQ Sequence 13 AA;  
Query Match 100.0%; Score 65; DB 22; Length 13;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VNWGVLAACAPSK 13  
Db 1 VNWGVLAACAPSK 13  
RESULT 3  
AAU07397 standard; protein: 13 AA.  
ID AAU07397 standard; protein: 13 AA.  
XX  
AC AAU07397;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Bacillus subtilis xylanase inhibitor #6.  
XX  
KM Xylanase; plant cell wall; baking; cereal; starch production; wood;  
KM wood pulp bleaching; hemicellulose; arabinoxylan; foodstuff; inhibitor.  
XX  
OS Bacillus subtilis.  
XX  
PN WO200166711-A1.  
XX  
PD 13-SEP-2001.  
XX  
PF 08-MAR-2001; 2001WO-1B00426.  
XX  
PR 08-MAR-2000; 2000GB-0005585.  
XX  
PR 27-JUN-2000; 2000GB-0015751.  
XX  
PA (DANI-) DANISCO AS.  
XX  
PI Sibsen O, Sorensen JF;  
XX  
DR WPI; 2001-596834/67.  
XX  
PT Novel variant xylanase polypeptide or its fragment useful for degrading  
PT or modifying plant cell wall, comprises amino acid modifications such  
PT that the polypeptide has altered sensitivity to xylanase inhibitor -  
XX  
PS Disclosure; Page 63; 70pp; English.  
XX  
CC The invention relates to a variant xylanase polypeptide (I) or its  
CC fragment having xylanase activity, comprising one or more amino acid  
CC modifications such that (I) or its fragment has an altered sensitivity to  
CC a xylanase inhibitor as compared with the parent xylanase enzyme. (I) or  
CC its coding sequence (II) is useful for degrading or modifying plant cell  
CC wall or for processing a plant material by contacting the plant cell wall  
CC or plant material with (I) or (II). (I) is useful for modifying plant  
CC materials, and in baking, processing cereals, starch production,  
CC processing wood and enhancing the bleaching of wood pulp. (I)  
CC is useful for altering the viscosity derived from the presence of  
CC hemicellulose or arabinoxylan in a solution or system comprising plant  
CC cell wall material. (I) is useful for preparing a foodstuff such as  
CC bread, pretzels, tortillas, cakes, cookies, biscuits or crackers.  
CC The present sequence represents the amino acid sequence of Bacillus  
CC subtilis xylanase inhibitor #6 as described in the method of the  
XX  
SQ Sequence 13 AA;  
Query Match 100.0%; Score 65; DB 22; Length 13;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNWGYLAACAPSK 13  
 |||  
 Db 1 VNWGYLAACAPSK 13

RESULT 4  
 AAU75815  
 ID AAU75815 standard; Protein: 370 AA.  
 XX AAU75815;  
 AC  
 XX  
 XX 23-APR-2002 (first entry)  
 DT  
 XX  
 DE Wheat L endoxylanase inhibitor, TAXI I, partial sequence TAXI-I.0.  
 KW Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruupng; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Triticum aestivum cultivar Soissons.  
 XX  
 PN WO200198474-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 XX 21-JUN-2001; 2001WO-BE00106.  
 PF  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.  
 XX  
 PI Delcours J, Debysier W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR WPI; 2002-114579/15.  
 DR N-PSDB; ABK13672.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprising using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 57; 127pp; English.

The invention relates to separating and/or isolating inhibitors of  
 cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 screening the inhibition activity by using two or more enzymes during the  
 separation and/or isolation steps that allow to distinguish inhibitors of  
 different specificity or by using an affinity chromatographic step with  
 immobilised enzymes and/or antibodies against inhibitors. Also  
 included are an isolated nucleic acid molecule encoding an inhibitor  
 which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 alpha-L-arabinofuranosidase and/or other cellulose, xylan,  
 arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 nucleic acid, an expression system transformed with the nucleic  
 acid, a host organism transformed with the nucleic acid, the inhibitory  
 proteins encoded by the nucleic acids and modulators of the proteins.  
 A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 plant or plant materials transformed with the nucleic acid are useful  
 for the formation of an endoxylanase-inhibitor complex, screening  
 endoxylanases that are totally, less or not inhibited by the inhibitors,  
 reducing syruupng in refrigerated dough compositions, affecting the  
 relative affinity and/or relative hydrolysis specificity and/or relative  
 hydrolysis rate versus water-extractable and/or water-unextractable

CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial wheat TAXI-I (T. aestivum L endoxylanase  
 CC inhibitor).  
 CC  
 XX  
 XX SQ Sequence 370 AA;  
 DE  
 DE Query Match 100.0%; Score 65; DB 23; Length 370;  
 DE Best Local Similarity 100.0%; Pred. No. 0.00093;  
 DE Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNWGYLAACAPSK 13  
 |||  
 Db 105 VNWGYLAACAPSK 117

RESULT 5  
 AAU75816  
 ID AAU75816 standard; Protein: 370 AA.  
 XX  
 AC AAU75816;  
 XX  
 XX 23-APR-2002 (first entry)  
 DT  
 XX  
 DE Wheat L endoxylanase inhibitor, TAXI I, partial sequence TAXI-I.02.  
 KW Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruupng; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Triticum aestivum cultivar Estica.  
 XX  
 PN WO200198474-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 XX 21-JUN-2001; 2001WO-BE00106.  
 PF  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.  
 XX  
 PI Delcours J, Debysier W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR WPI; 2002-114579/15.  
 DR N-PSDB; ABK13673.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprising using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 57; 127pp; English.

The invention relates to separating and/or isolating inhibitors of  
 cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 screening the inhibition activity by using two or more enzymes during the

separation and/or isolation steps that allow to distinguish inhibitors of different specificity or by using an affinity chromatographic step with immobilised enzymes and/or antibodies against inhibitors. Also included are an isolated nucleic acid molecule encoding an inhibitor which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase, alpha-L-arabino-furanosidase and/or other cellulose, xylan, arabinoxylan or beta-glucan degrading enzymes, a vector comprising the nucleic acid, an expression system transformed with the nucleic acid, a host organism transformed with the nucleic acid, the inhibitory proteins encoded by the nucleic acids and modulators of the proteins. A recombinant protein, glycoprotein or polypeptide or microorganisms, plant or plant materials transformed with the nucleic acid are useful for the formation of an endoxylanase-inhibitor complex. Screening endoxylanases that are totally, less or not inhibited by the inhibitors, reducing syripping in refrigerated dough compositions, affecting the relative affinity and/or relative hydrolysis specificity and/or relative hydrolysis rate versus water-extractable and/or water-unextractable arabinoxylans of endoxylanases such as by the formation of an endoxylanase/inhibitor complex, improving the malting of cereals such as barley, sorghum and wheat and/or the production of beer, improving the production and/or quality of baked or extruded cereal products such as straight dough, sponge dough, Chorleywood bread, breakfast cereals, biscuits, pasta and noodles, animal feed stuff, improving the production of starch derived syrups, sorbitol, xylitol, xylose and/or xylitol, wheat gluten starch separation and production, improving maize processing, plant disease resistance and nutraceutical and/or pharmaceutical applications, improving paper and pulp technologies. The present sequence represents a partial wheat TAXI-I (T. aestivum L endoxylanase inhibitor).

SO Sequence 370 AA:

Query Match 100.0%; Score 65; DB 23; Length 370;  
Best Local Similarity 100.0%; Pred. NO. 0.00095;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNWGYLAACPSK 13

DB 105 VNWGYLAACPSK 117

RESULT 6  
AAU75812  
ID AAU75812 standard; Protein: 381 AA.

AC AAU75812;

DT 23-APR-2002 (first entry)

DE Wheat L endoxylanase inhibitor; TAXI I; variant #1.

XX wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
XX xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
XX immobilised enzyme; enzyme; dough syripping; cereal product; beer; plant;  
XX straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
XX food; animal feed; starch separation; maize processing; malting;  
XX plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
XX Triticum aestivum.

OS Triticum aestivum.

FT key Location/Qualifiers

FT MISC-difference 95 /Label= Unknown

FT MISC-difference 98 /Label= Unknown

FT MISC-difference 101 /Label= Unknown

FT MISC-difference 110 /Label= Unknown

FT MISC-difference 333 /Label= Unknown

PN WO200198474-A1.

XX 27-DEC-2001.

PD 21-JUN-2001; 2001WO-BE00106.

PR 22-JUN-2000; 2000GB-0015296.

PR 25-JAN-2001; 2001GB-0002018.

PR 26-JAN-2001; 2001GB-0002194.

PR 16-MAR-2001; 2001GB-0006564.

PR 21-MAY-2001; 2001GB-0012328.

XX (LEUV-) LEUVEN RES & DEV.

XX Delcourt J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J;

PI Van Campenhout S;

PS WPI; 2002-114579/15.

PS Claim 127; Page 50-51; 127pp: English.

CC The invention relates to separating and/or isolating inhibitors of

CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises

CC screening the inhibition activity by using two or more enzymes during the

CC separation and/or isolation steps that allow to distinguish inhibitors of

CC different specificity or by using an affinity chromatographic step with

CC immobilised enzymes and/or antibodies against inhibitors. Also

CC included are an isolated nucleic acid molecule encoding an inhibitor

CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,

CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,

CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the

CC nucleic acid, an expression system transformed with the nucleic

CC acid, a host organism transformed with the nucleic acid, the inhibitory

CC proteins encoded by the nucleic acids and modulators of the proteins.

CC A recombinant protein, glycoprotein or polypeptide or microorganisms,

CC plant or plant materials transformed with the nucleic acid are useful

CC for the formation of an endoxylanase-inhibitor complex, screening

CC endoxylanases that are totally, less or not inhibited by the inhibitors,

CC reducing syripping in refrigerated dough compositions, affecting the

CC relative affinity and/or relative hydrolysis specificity and/or relative

CC hydrolysis rate versus water-extractable and/or water-unextractable

CC arabinoxylans of endoxylanases such as by the formation of an

CC endoxylanase/inhibitor complex, improving the malting of cereals such as

CC barley, sorghum and wheat and/or the production of beer, improving the

CC production and/or quality of baked or extruded cereal products such as

CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,

CC biscuits, pasta and noodles, animal feed stuff, improving the production

CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten

CC starch separation and production, improving maize processing,

CC plant disease resistance and nutraceutical and/or pharmaceutical

CC applications, improving paper and pulp technologies. The present

CC sequence is a variant of wheat TAXI-I (T. aestivum L endoxylanase

CC inhibitor).

CC Note: Variant amino acids are highlighted in the specification but

CC no wild-type sequence is shown for comparison.

SO Sequence 381 AA:

Query Match 100.0%; Score 65; DB 23; Length 381;  
Best Local Similarity 100.0%; Pred. NO. 0.00098;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNWGYLAACPSK 13

DB 116 VNWGYLAACPSK 128

RESULT 7

AAU75813

ID	AAU75813 standard; Protein; 381 AA.
XX	
AC	AAU75813;
XX	
DT	23-APR-2002 (first entry)
DE	
XX	
XX	Wheat L endoxylanase inhibitor, TAXI I, variant #2.
KW	Wheat; TAXI-I: L endoxylanase inhibitor; cellulolytic enzyme inhibitor;
KW	xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;
KW	straight dough; sponge dough; Chorleywood bread; biscuits; pasta;
KW	noodle; animal feed; starch separation; maize processing; malting;
KW	plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.
XX	
OS	Triticum aestivum.
XX	
FT	Key Location/Qualifiers
FT	Misc-difference 95
FT	/label= Unknown
FT	Misc-difference 98
FT	/label= Unknown
FT	Misc-difference 101
FT	/label= Unknown
FT	Misc-difference 110
FT	/label= Unknown
FT	Misc-difference 333
FT	/label= Unknown
XX	
PM	WO200198474-A1.
XX	
PD	27-DEC-2001.
XX	
PF	21-JUN-2001; 2001WO-BE00106.
XX	
PR	22-JUN-2000; 2000GB-0015296.
PR	25-JAN-2001; 2001GB-0002018.
PR	26-JAN-2001; 2001GB-0002194.
PR	16-MAR-2001; 2001GB-0006564.
PR	21-MAY-2001; 2001GB-0012328.
XX	
PA	(LEUV-) LEUVEN RES & DEV.
XX	
PI	Deleour J, Debyser W, Gebruers K, Goesaert H, Fiersens K, Robben J;
PI	Van Campenhout S;
XX	
DR	WPI. 2002-114579/15.
XX	
PT	Separating and/or isolating inhibitors of cellulolytic, xyylanolytic, or
PT	beta-glucanolytic enzymes comprises using endoxylanases during
PT	screening for inhibition activity or affinity chromatography with
PT	immobilised enzymes
XX	
PS	Claim 127; Page 51; 127pp; English.
XX	
CC	The invention relates to separating and/or isolating inhibitors of
CC	cellulolytic, xyylanolytic and/or beta-glucanolytic enzymes comprises
CC	screening the inhibition activity by using two or more enzymes during the
CC	separation and/or isolation steps that allow to distinguish inhibitors of
CC	different specificity or by using an affinity chromatographic step with
CC	immobilised enzymes and/or antibodies against inhibitors. Also
CC	included are an isolated nucleic acid molecule encoding an inhibitor
CC	which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,
CC	alpha-L-arabino-furanosidase and/or other cellulose, xylan,
CC	arabinoxylan or beta-glucan degrading enzymes, a vector comprising the
CC	nucleic acid, an expression system transformed with the nucleic
CC	acid, a host organism transformed with the nucleic acid, the inhibitory
CC	proteins encoded by the nucleic acids and modulators of the proteins.
CC	A recombinant protein, glycoprotein or polypeptide or microorganisms,
CC	plant or plant materials transformed with the nucleic acid are useful
CC	for the formation of an endoxylanase-inhibitor complex, screening
CC	endoxylanases that are totally, less or not inhibited by the inhibitors,
CC	reducing syringing in retrefrigerated dough compositions, affecting the

CC	relative affinity and/or relative hydrolysis specificity and/or relative
CC	hydrolysis rate versus water-extractable and/or water-unextractable
CC	arabinosylans of endoxylanases such as by the formation of an
CC	endoxylanase/inhibitor complex, improving the malting of cereals such as
CC	barley, sorghum and wheat and/or the production of beer, improving the
CC	production and/or quality of baked or extruded cereal products such as
CC	straight dough, sponge dough, Chorleywood bread, breakfast cereals,
CC	biscuits, pasta and noodles, animal feed stuff, improving the production
CC	of starch derived syrups, sorbitol, xylitol, xylose and/or xylitol, wheat gluten
CC	starch separation and production, improving maize processing,
CC	plant disease resistance and nutraceutical and/or pharmaceutical
CC	applications, improving paper and pulp technologies. The present
CC	sequence is a variant of wheat TAXI-1 ( <i>T. aestivum</i> L endoxylanase
CC	inhibitor).
CC	Note: Variant amino acids are highlighted in the specification but
CC	no wild-type sequence is shown for comparison.
XX	
SO	Sequence      381 AA;
OY	Query Match                  100.0%; Score 65; DB 23; Length 381; Best Local Similarity        100.0%; Pred. No. 0.00098; Matches    13; Conservative    0; Mismatches    0; Indels     0; Gaps     0;
ID	1 VNWGYLACAPSK 13       Db        116 VNWGYLACAPSK 128
RESUL.T 8	
AAU75814	AAU75814 standard; Protein: 381 AA.
XX	AAU75814;
DT	23-APR-2002 (first entry)
XX	
DE	Wheat L endoxylanase inhibitor, TAXI 1, microheterogenic variant.
XX	
KW	Wheat; TAXI-i: L endoxylanase inhibitor; cellulolytic enzyme inhibitor;
KW	xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;
KW	immobilised enzyme; enzyme; dough syruupng; cereal product; beer; plant;
KW	straight dough; sponge dough; Chorleywood bread; biscuits; pasta;
KW	noodle; animal feed; starch separation; maize processing; malting;
KW	plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.
XX	
OS	Triticum aestivum.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 95 /label= Unknown
FT	Misc-difference 98 /label= Unknown
FT	Misc-difference 101 /label= Unknown
FT	Misc-difference 110 /label= Unknown
FT	Misc-difference 145 /label= Unknown
FT	Misc-difference 183 /label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 232 /label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 275 /label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 282 /label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
FT	Misc-difference 333 /label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys
XX	/label= Unknown
PN	WO200198474-A1.
XX	
PD	27-DEC-2001.
XX	

PF 21-JUN-2001; 2001WO-BE00106.  
 XX  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 XX (LEUV-) LEUVEN RES & DEV.  
 PA  
 PI Delcour J, Debyser W, Gebruers K, Goesaert H, Flereens K, Robben J;  
 PI Van Campenhout S;  
 DR MPI; 2002-114579/15.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 51; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a microheterogenic variant of wheat TAXI-I (T. aestivum L  
 CC endoxylanase inhibitor).  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.  
 XX  
 SO Sequence 381 AA:  
 Query Match 100.0%; Score 65; DB 23; Length 381;  
 Best Local Similarity 100.0%; Pred. No. 0.00098;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNVGVLAACAPSK 13  
 |||||||||  
 Db 116 VNVGVLAACAPSK 128

RESULT 9  
 AAU75819  
 ID AAU75819 standard; Protein; 287 AA.  
 XX  
 AC AAU75819;

XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Wheat L endoxylanase inhibitor, partial sequence TAXI-III.  
 XX  
 KW wheat; TAXI-III; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Triticum aestivum cultivar Soissons.  
 XX  
 PN WO200198474-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 21-JUN-2001; 2001WO-BE00106.  
 XX  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 PA (LEUV-) LEUVEN RES & DEV.  
 XX  
 PI Delcour J, Debyser W, Gebruers K, Goesaert H, Flereens K, Robben J;  
 PI Van Campenhout S;  
 DR MPI; 2002-114579/15.  
 XX  
 DR N-PSDB; ABK13676.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 58; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial wheat TAXI-III (T. aestivum L endoxylanase  
 CC inhibitor).



XX SQ Sequence 287 AA: 76.9%; Score 50; DB 23; Length 287;  
 Query Match Best Local Similarity 76.9%; Pred. No. 0.43;  
 Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VNWGVLAACAPSK 13  
 |||:|||||||  
 Db 107 VNVRAVMAACAPSK 119

RESULT 10  
 AAU75818  
 ID AAU75818 standard; Protein; 277 AA.  
 XX AAU75818;  
 AC  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE L endoxylanase inhibitor, ATXI-II, partial sequence ATXI-II.01.  
 XX  
 KW ATXI-II; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor; ds;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Aegilops tauschii variety Acc2220051.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 4 /note- "Encoded by CTC"  
 XX  
 XX WO200198474-A1.  
 XX  
 XX 27-DEC-2001.  
 XX  
 PD 21-JUN-2001; 2001WO-BE00106.  
 XX  
 PF 22-JUN-2000; 2000GB-0015296.  
 XX  
 PR 25-JAN-2001; 2001GB-0002018.  
 XX  
 PR 26-JAN-2001; 2001GB-0002194.  
 XX  
 PR 16-MAR-2001; 2001GB-0006564.  
 XX  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 XX (LEUV-) LEUVEN RES & DEV.  
 PA  
 XX Delcour J, Debysse W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 XX  
 DR WPI: 2002-114579/15.  
 XX  
 DR N-PSDB: ABK13675.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 58; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-galcan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic

CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial Aegilops tauschii ATXI-II (Aegilops  
 CC tauschii L endoxylanase inhibitor).

XX SQ Sequence 277 AA: 66.2%; Score 43; DB 23; Length 277;  
 Query Match Best Local Similarity 66.2%; Pred. No. 8.3;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 VNWGVLAACAPSK 13  
 |||:|||||||  
 Db 105 VVVRILAACAPRK 117

RESULT 11  
 AAW81605  
 ID AAW81605 standard; peptide; 25 AA.  
 XX  
 AC AAW81605;  
 XX  
 DT 27-JAN-1999 (first entry)  
 XX  
 DE Synthetic rabies nucleoprotein homologue peptide fragment.  
 XX  
 XX 9D: vaccine; cytotoxic T cell response; viral infection; rabies;  
 KW peptide-fatty acid conjugate; influenza; oncogene; liposome; APC;  
 KW human immune deficiency virus; antigen-presenting cell; antibody;  
 KW protection.  
 XX  
 OS Synthetic.  
 XX  
 PN US5837249-A.  
 XX  
 PD 17-NOV-1998.  
 XX  
 PD 20-OCT-1993; 93US-0139609.  
 XX  
 PF 15-APR-1992; 92US-0868946.  
 XX  
 PR 19-APR-1985; 85US-0725087.  
 XX  
 PR 08-MAY-1987; 87US-0047443.  
 XX  
 PR 12-APR-1991; 91US-0685459.  
 XX  
 PR 20-OCT-1993; 93US-0139609.  
 XX  
 PA (WIST-) WISTAR INST.  
 XX  
 XX Dietzschold B, Heber-Katz E;  
 PI  
 XX  
 DR WPI: 1999-023378/02.  
 XX  
 PT Inducing cytotoxic T cell response against virus using peptide-fatty  
 PT acid conjugate - Formulated in liposomes with an adjuvant,  
 PT specifically for protecting against herpes simplex or rabies viruses  
 XX  
 PS Example 11; Column 14; 21pp; English.

XX This sequence is a synthetic peptide fragment which is a homologue of the  
 CC rabies virus N protein. This sequence is used in a method of inducing  
 CC a cytotoxic T cell response against viral infection in a mammal by  
 CC administering a peptide-fatty acid conjugate. Such conjugates are used  
 CC particularly to vaccinate against HSV, rabies and also other viruses such  
 CC as influenza, human immune deficiency virus and oncogenic viruses. When  
 CC the liposomes fuse to an antigen-presenting cell (APC), the conjugate  
 CC remains bound to the surface of the APC membrane and is not degraded  
 CC inside the cell, which generates a T cell response without any antibody  
 CC response, avoiding the risk of immune enhancement (in which antibodies  
 CC increase viral infectivity). Such compounds can provide long-lasting  
 CC protection from only a single injection.

CC Sequence 25 AA:

Query Match 63.1%; Score 41; DB 20; Length 25;  
 Best Local Similarity 72.7%; Pred. No. 1.4;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VNVGLAACAP 11  
 : | : |||||  
 DB 13 LNATVIAACAP 23

RESULT 12

AAP80639  
 ID AAP80689 standard; peptide; 25 AA.

AC AAP80689;

DT 07-NOV-1990 (first entry)

DE Sequence corresp. to a peptide fragment derived from rabies  
 DE virus nucleoprotein.

XX Immunodominant T-cell antigen; rabies vaccines;  
 KW peptide-fatty acid conjugate.

OS Rabies virus.

PN EP290246-A.

PD 09-NOV-1988.

PF 05-MAY-1988; 88EP-0304045.

PR 08-MAY-1987; 87US-0047443.

PA (WIST-) WISTAR INST ANATOMY.

PI Heber-Katz E, Dietzschold B;

DR WPI; 1988-316484/45.

PT Peptide-fatty acid conjugates;  
 used in vaccines for generating an immunogenic T cell response  
 PT protective against a virus disease state e.g. rabies

PS Claim 6; Page 22; 36pp; English.

CC Peptide fragment is an immunodominant T cell antigenic determinant. It  
 CC is part of a new peptide-fatty acid conjugate. Also claimed is a vaccine  
 CC for generating an immunogenic T cell response protective against a virus  
 CC disease state comprising:  
 CC (a) a peptide-fatty acid complex or a synthetic replica;  
 CC (b) a liposome compsn. comprising a mixt. of phosphatidyl choline,  
 CC cholesterol and lysophosphatidyl cholines, (proportions by wt. pref.  
 CC 16:2:1); and  
 CC (c) an adjuvant.

CC Vaccines achieve significant protection for a protracted period of time  
 CC against a large dose of virus by a single injection. A peptide having  
 CC sequence homology to the 25 AA peptide chain of rabies nucleoprotein,

CC having a spacer of Gly-Gly-Lys added to N-terminus and Cys added to the  
 CC C-terminus and having the formula in AAP80689 was prepd. One month after  
 CC a single immunisation (10ug peptide/animal), mice were challenged with a  
 CC lethal dose of rabies virus (2 LD<sub>50</sub>) dose of strain CVS-24 grown in BHK  
 CC cells) and the animals examined for the next 21 days for paralysis and  
 CC death; 87% of vaccine injected animals survived.

CC Sequence 25 AA:

Query Match 60.0%; Score 39; DB 9; Length 25;  
 Best Local Similarity 63.6%; Pred. No. 3.2;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VNVGLAACAP 11  
 : | : |||||  
 DB 13 LNATVIAACAP 23

RESULT 13

AAP91478  
 ID AAP91478 standard; protein; 25 AA.

AC AAP91478;

DT 24-JUN-1990 (first entry)

DE Peptide N-V12b.

KW Rabies related virus; peptide N-V12b; vaccine.

OS Rabies related virus.

PN WO8900861-A.

PD 09-FEB-1989.

PF 29-JUL-1988; 88WO-US02529.

PR 30-JUL-1987; 87US-0079639.

PA (WIST-) THE WISTAR INSTITUTE.

PI Dietzschold B, Koprowski H;

DR WPI; 1989-061041/08.

PT New rabies virus polypeptide(s) - useful for vaccine prodn.

PS Claim 13; page 25; 31pp; English.

CC Peptide N-V10b corresponds to amino acids 313-337 of rabies related virus  
 CC (including rabies virus and related viruses such as Mokola and European  
 CC bat virus) nucleoprotein. It is useful for the prodn of vaccines against  
 CC homologous and heterologous rabies virus strains. They can be produced  
 CC without growing the virus, and do not cause the adverse side effects of  
 CC inactivated virus vaccines, eg demyelinating allergic encephalitis and  
 CC systemic reactions. The peptide is opt. administered in liposomes  
 CC (coupled to 15-21C fatty acids beforehand), followed by at least one  
 CC booster injection, esp. of inactivated virus or a mixt. of virus  
 CC glycoprotein and nucleoprotein.

CC Sequence 25 AA:

Query Match 60.0%; Score 39; DB 10; Length 25;  
 Best Local Similarity 63.6%; Pred. No. 3.2;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VNVGLAACAP 11  
 : | : |||||  
 DB 13 LNATVIAACAP 23

RESULT 14

AAW81603  
ID AAW81603 standard; peptide; 25 AA.  
XX  
AC AAW81603;  
DT 27-JAN-1999 (first entry)  
XX  
DE Synthetic peptide fragment N-VI2b.  
XX  
KM gD; vaccine; cytotoxic T cell response; viral infection; rabies;  
KM peptide-fatty acid conjugate; influenza; oncogene; liposome; APC;  
KM human immune deficiency virus; antigen-presenting cell; antibody;  
KM protection.  
XX  
OS Synthetic.  
XX  
PN US837249-A.  
PD 17-NOV-1998.  
XX  
XX 20-OCT-1993; 93US-0139609.  
PF  
XX 15-APR-1992; 92US-0868946.  
PR 19-APR-1985; 85US-0725087.  
PR 08-MAY-1987; 87US-0047443.  
PR 12-APR-1991; 91US-0685459.  
PR 20-OCT-1993; 93US-0139609.  
XX  
PA (WIST-) WISTAR INST.  
XX  
PI Dietzschold B, Heber-Katz E;  
XX  
DR WPI; 1999-023378/02.  
XX  
PT Inducing cytotoxic T cell response against virus using peptide-fatty  
PT acid conjugate - formulated in liposomes with an adjuvant,  
PT specifically for protecting against herpes simplex or rabies viruses  
XX  
PS Claim 21; Column 4; 21pp; English.  
XX  
XX This sequence is a synthetic peptide fragment, N-VI2b, which is a  
CC homologue of the rabies virus N protein. This sequence is used in a  
CC method of inducing a cytotoxic T cell response against viral infection in  
CC a mammal by administering a peptide-fatty acid conjugate. Such conjugates  
CC are used particularly to vaccinate against HSV, rabies and also other  
CC viruses such as influenza, human immune deficiency virus and oncogenic  
CC viruses. When the liposomes fuse to an antigen-presenting cell (APC), the  
CC conjugate remains bound to the surface of the APC membrane and is not  
CC degraded inside the cell, which generates a T cell response without any  
CC antibody response, avoiding the risk of immune enhancement (in which  
CC antibodies increase viral infectivity). Such compounds can provide  
CC long-lasting protection from only a single injection.  
XX  
SQ Sequence 25 AA;  
XX  
Query Match 60.0%; Score 39; DB 20; Length 25;  
Best Local Similarity 63.6%; Pred. No. 3.2;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 VNVGVLAACAP 11  
! !:|||||  
Db 13 LNMATVIACAP 23  
XX  
RESULT 15  
AAG42355  
ID AAG42355 standard; Protein; 341 AA.  
XX  
AC AAG42355;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52814.

XX  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142877.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 15-JUL-1999; 99US-0144334.  
PR 15-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145092.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 25-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 23-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158362.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 60.0%; Score 39; DB 21; Length 341;  
Best Local Similarity 50.0%; Pred. No. 58;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 2 NNGVLAACPSK 13  
DB 3 SIVGTALCSPAK 14

Search completed: May 20, 2003, 16:14:43  
Job time : 15.1429 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:11:22 ; Search time 5.28571 seconds  
(without alignments)  
236.439 Million cell updates/sec

Title: US-09-869-155-18  
Perfect score: 65  
Sequence: 1 VNVGVLAACAPSK 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	63.1	169	2 C75427	hypothetical prote
2	40	61.5	228	2 F75390	hypothetical prote
3	40	61.5	349	2 F95263	selenide, water di
4	40	61.5	493	2 S11148	ama protein - str
5	40	61.5	659	2 A95221	hypothetical prote
6	40	61.5	659	2 A98085	hypothetical prote
7	39	60.0	257	2 AC2406	2-hydroxyhepta-2,4
8	39	60.0	286	2 T24066	exonuclease III ho
9	39	60.0	290	2 T09363	hypothetical prote
10	39	60.0	409	2 S12018	endoglucanase B -
11	39	60.0	437	2 JO2015	nucleoprotein - ra
12	39	60.0	437	2 JO2016	nucleoprotein - ra
13	39	60.0	450	1 VHNVRY	nucleoprotein - ra
14	39	60.0	450	1 VHNVNS	nucleoprotein - ra
15	39	60.0	450	1 VHNVAV	nucleoprotein - ra
16	38	58.5	131	2 A87461	hypothetical prote
17	38	58.5	208	2 F70734	hypothetical prote
18	38	58.5	384	2 AH0241	ceramide glucosylt
19	38	58.5	835	2 C86444	hypothetical prote
20	37	56.9	147	2 S77485	ribosomal protein
21	37	56.9	263	2 D71281	probable pyroline
22	37	56.9	334	2 D87575	hypothetical prote
23	37	56.9	409	2 F70633	probable lipk prot
24	37	56.9	431	2 T50573	probable integral
25	37	56.9	450	2 A58460	nucleoprotein N -
26	37	56.9	459	2 D70914	hypothetical prote
27	37	56.9	1117	2 A38227	RNA-splicing regul
28	36	55.4	155	2 E5174	hypothetical 16.7
29	36	55.4	247	2 T35602	probable transpos

30	36	55.4	359	2 A13259	hypothetical prote
31	36	55.4	385	2 G71474	probable cell divi
32	36	55.4	396	2 A13554	xylose transport s
33	36	55.4	407	2 H81736	cell division prot
34	36	55.4	471	2 D83388	probable MFS trans
35	36	55.4	559	2 H84859	hypothetical prote
36	36	55.4	601	2 B96744	unknown protein [i
37	36	55.4	738	2 C84700	hypothetical prote
38	35.5	54.6	621	2 H87652	feruloyl-coa synth
39	35	53.8	128	2 T28433	selenophosphate sy
40	35	53.8	179	2 S63610	probable prophage
41	35	53.8	179	2 JC6047	regB protein - Ser
42	35	53.8	182	2 E82220	conserved hypothet
43	35	53.8	210	2 C87256	hypothetical prote
44	35	53.8	217	2 S62795	probable lipoprote
45	35	53.8	278	2 G81139	hypothetical prote

## ALIGNMENTS

```

RESULT 1
C75427
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
C:Accession: C75427
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: C75427
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-169 <WHI>
A:Map position: 1
A:Gene: DR1180
A:Superfamily: Deinococcus radiodurans hypothetical protein DR1180
C:Superfamily: Deinococcus radiodurans hypothetical protein DR1180

Query Match          63.1%  Score 41;  DB 2;  Length 169;
Best Local Similarity 70.0%  Pred. No. 6.9;
Matches 7;  Conservative 3;  Mismatches 0;  Indels 0;  Gaps 0;

OY      3  VGVLAACAPS 12
DB      13  VGLVLAACAP 22

RESULT 2
F75390
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75390
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75390
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-228 <WHI>
A:Cross-references: GB:AE001992; GB:AE000513; NID:g6459233; PIDN:AAF11048.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1474

```

A:Map position: 1

Query Match 61.5%; Score 40; DB 2; Length 228;  
 Best Local Similarity 77.8%; Pred. No. 13;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 GVLAACAP 11  
 11:11:111  
 DB 14 GVLAACAP 22

RESULT 3

F95263

selenide, water dikinase (EC 2.7.9.3) seld [imported] - Shnorhizobium meliloti (strain 1  
 N:Alternate names: selenophosphate synthetase

C:Species: Shnorhizobium meliloti

C&gt;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001

C:Accession: F95263

R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows  
 ; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001

A:Title: Nucleotide sequence and predicted functions of the entire Shnorhizobium meliloti

A:Reference number: A95262; MUID:21396509; PMID:11481432

A:Accession: F95263

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 &lt;KUR&gt;

A:Cross-references: GB:AE006469; PIDN:AAK64672.1; PID:g14523070; GSPDB:GN00165

A:Experimental source: strain 1021, megaplasmid pSymA

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 P.; D., Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 hebbault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont Shnorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Accession: A96039

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 &lt;KUR&gt;

A:Cross-references: GB:AE006469; PIDN:AAK64672.1; PID:g14523070; GSPDB:GN00165

A:Experimental source: strain 1021, megaplasmid pSymA

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 P.; D., Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 hebbault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont Shnorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Accession: A96039

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 &lt;KUR&gt;

A:Cross-references: GB:AE006469; PIDN:AAK64672.1; PID:g14523070; GSPDB:GN00165

A:Experimental source: strain 1021, megaplasmid pSymA

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 P.; D., Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 hebbault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont Shnorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Accession: A96039

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 &lt;KUR&gt;

A:Cross-references: GB:AE006469; PIDN:AAK64672.1; PID:g14523070; GSPDB:GN00165

A:Experimental source: strain 1021, megaplasmid pSymA

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 P.; D., Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 hebbault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont Shnorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Accession: A96039

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 &lt;KUR&gt;

A:Cross-references: GB:AE006469; PIDN:AAK64672.1; PID:g14523070; GSPDB:GN00165

A:Experimental source: strain 1021, megaplasmid pSymA

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 P.; D., Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

DB 18 GVLAACSSK 27  
 111111: 11

Query Match 61.5%; Score 40; DB 2; Length 659;  
 Best Local Similarity 80.0%; Pred. No. 33;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 GVLAACAPSK 13  
 111111: 11  
 DB 18 GVLAACSSK 27

RESULT 5

A95221

hypothetical protein SP1891 [imported] - Streptococcus pneumoniae (strain TIGR4)  
 C:Species: Streptococcus pneumoniae

C&gt;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: A95221

R:Rettelli, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H  
 on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfl  
 son, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11465916

A:Accession: A95221

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-659 &lt;KUR&gt;

A:Cross-references: GB:AE005672; PIDN:AAK75962.1; PID:g14973394; GSPDB:GN00164; TIGR  
 A:Experimental source: strain TIGR4

C:Genetics: SP1891

Query Match 61.5%; Score 40; DB 2; Length 659;

Best Local Similarity 80.0%; Pred. No. 33;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 GVLAACAPSK 13

111111: 11

DB 18 GVLAACSSK 27

RESULT 6

A98085

hypothetical protein amia [imported] - Streptococcus pneumoniae (strain R6)  
 C:Species: Streptococcus pneumoniae

C&gt;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C:Accession: A98085

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Letkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
 y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: A98085

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-659 &lt;KUR&gt;

A:Cross-references: GB:AE007317; PIDN:AAU00510.1; PID:g15459383; GSPDB:GN00174

C:Genetics: amia

Query Match 61.5%; Score 40; DB 2; Length 659;

Best Local Similarity 80.0%; Pred. No. 33;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 GVLAACAPSK 13

111111: 11

DB 18 GVLAACSSK 27

RESULT 7

AC2406

2-hydroxyhepta-2,4-diene-1,7-dioate isomerase [imported] - Nostoc sp. (strain PCC 712  
 C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C&gt;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AC2406

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu

Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AC2406  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-257 <KUR>  
 A:Cross-References: GB:BA000019; PIDN:BAW76502.1; PID:g17133940; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: alr4803  
 C:Superfamily: 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase

Query Match 60.0%; Score 39; DB 2; Length 257;  
 Best Local Similarity 77.8%; Pred. No. 22;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 5 VLAACAPSK 13  
 :|||  
 Db 51 ILAPCAPSK 59

RESULT 8  
 T24066  
 exonuclease III homolog - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T24066; T42391  
 R:Kershaw, J.  
 submitted to the EMBL Data Library, October 1996  
 A:Reference number: Z19837  
 A:Accession: T24066  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-288 <MIL>  
 A:Cross-References: EMBL:281108; PIDN:CA03235.1; GSPDB:GN00019; CESP:R09B3.1  
 A:Experimental source: clone R09B3  
 R:Tremblay, S.; Masson, J.Y.; Ramotar, D.  
 submitted to the EMBL Data Library, November 1997  
 A:Description: The exonuclease III family of DNA repair enzymes has a new homolog in *Cer*  
 A:Reference number: Z22150  
 A:Accession: T42391  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-288 <PRE>  
 A:Cross-References: EMBL:AF034258; PIDN:AAC82328.1  
 C:Genetics:  
 A:Gene: R09B3.1  
 A:Map position: 1  
 A:Introns: 83/3; 192/3; 264/3  
 C:Superfamily: exodeoxyribonuclease III

Query Match 60.0%; Score 39; DB 2; Length 288;  
 Best Local Similarity 63.6%; Pred. No. 24;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 VGLVLAACAPSK 13  
 :|||  
 Db 102 VGLLSKCAPSK 112

RESULT 9  
 T09363  
 hypothetical protein F23K16.100 - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 22-Oct-1999  
 C:Accession: T09363  
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.  
 submitted to the Protein Sequence Database, June 1999  
 A:Reference number: Z16652  
 A:Accession: T09363  
 A:Molecule type: DNA

A:Residues: 1-290 <BEV>  
 A:Cross-References: EMBL:AL078620; GSPDB:GN00062; ATSP:F23K16.100  
 A:Experimental source: cultivar Columbia; BAC clone F23K16  
 C:Genetics:  
 A:Gene: ATSP:F23K16.100  
 A:Map position: 4  
 A:Introns: 205/3; 253/3

Query Match 60.0%; Score 39; DB 2; Length 290;  
 Best Local Similarity 50.0%; Pred. No. 24;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 NVGVLAACAPSK 13  
 :|||  
 Db 3 SIGVTALCSPAK 14

RESULT 10  
 S12018  
 endoglucanase B - *Ruminococcus albus*  
 C:Species: *Ruminococcus albus*  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 29-Sep-1999  
 C:Accession: S12018  
 R:Poole, D.M.; Hazlewood, G.P.; Laurie, J.I.; Barker, P.J.; Gilbert, H.J.  
 Mol. Gen. Genet. 223, 217-223, 1990  
 A:Title: Nucleotide sequence of the *Ruminococcus albus* SY3 endoglucanase genes *celA*  
 A:Reference number: S12017; MUID:91066833; PMID:2250649  
 A:Accession: S12018  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-409 <POO>  
 A:Cross-References: GB:X54932; NID:g45965; PIDN:CAA38693.1; PID:g45966  
 C:Superfamily: *Ruminococcus albus* cellulase

Query Match 60.0%; Score 39; DB 2; Length 409;  
 Best Local Similarity 53.8%; Pred. No. 33;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 VNVGVLAACAPSK 13  
 :|||  
 Db 15 MSVGVMAACGCGSK 27

RESULT 11  
 J02015  
 nucleoprotein - rabies virus (isolate T2)  
 C:Species: rabies virus  
 C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 07-May-1999  
 C:Accession: J02015  
 R:Nadin-Davis, S.A.; Casey, G.A.; Wandeler, A.  
 J. Gen. Virol. 74, 829-837, 1993  
 A:Title: Identification of regional variants of the rabies virus within the Canadian  
 A:Reference number: J02012; MUID:93260396; PMID:8492088  
 A:Accession: J02015  
 A:Molecule type: genomic RNA  
 A:Residues: 1-437 <NAD>  
 C:Genetics:  
 A:Gene: N  
 C:Superfamily: rabies virus nucleoprotein  
 C:Keywords: nucleoprotein

Query Match 60.0%; Score 39; DB 2; Length 437;  
 Best Local Similarity 63.6%; Pred. No. 35;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VNVGVLAACAP 11  
 :|||  
 Db 320 LNAVTVLAACAP 330

RESULT 12  
 J02016  
 nucleoprotein - rabies virus (isolate T3)

C:Species: rabies virus  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 07-May-1999  
C:Accession: JQ2016  
R:Nadin-Davis, S.A.; Casey, G.A.; Wandeler, A.  
J.Gen.Virol. 74, 829-837, 1993  
A:Title: Identification of regional variants of the rabies virus within the Canadian pro  
A:Reference number: JQ2012; MUID:93260396; PMID:8492088  
A:Accession: JQ2016  
A:Molecule type: genomic RNA  
A:Residues: 1-437 <NAD>  
C:Genetics:  
A:Gene: N  
C:Superfamily: rabies virus nucleoprotein  
C:Keywords: nucleoprotein

Query Match 60.0%; Score 39; DB 2; Length 437;  
Best Local Similarity 63.6%; Pred. No. 35;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VNWGLAACAP 11  
: | | | | | | |  
DB 325 LNATVIAACAP 330

## RESULT 13

VHVNRY  
nucleoprotein - rabies virus  
C:Species: rabies virus  
C:Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 16-Jul-1999  
C:Accession: D26275; A24887  
R:Torco, N.; Poch, O.; Ermline, A.; Keith, G.; Rougeon, F.  
Proc. Natl. Acad. Sci. U.S.A. 83, 3914-3918, 1986  
A:Title: Walking along the rabies genome: is the large G-L intergenic region a remnant  
A:Reference number: A94100; MUID:86233343; PMID:3459163  
A:Accession: D26275  
A:Molecule type: genomic RNA  
A:Residues: 1-450 <TOR>  
A:Cross-references: GB:M13215; GB:M21634; NID:9333585; PIDN:AAA47215.1; PID:9333586  
C:Genetics:  
A:Gene: N  
C:Superfamily: rabies virus nucleoprotein  
C:Keywords: nucleoprotein

Query Match 60.0%; Score 39; DB 1; Length 450;  
Best Local Similarity 63.6%; Pred. No. 36;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VNWGLAACAP 11  
: | | | | | | |  
DB 325 LNATVIAACAP 335

## RESULT 14

VHVNRS  
nucleoprotein - rabies virus (strain SAD B19)  
C:Species: rabies virus  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Jul-1999  
C:Accession: A34746  
R:Conzelmann, K.K.; Cox, J.H.; Schneider, L.G.; Thiel, H.J.  
Virolgy 175, 485-499, 1990  
A:Title: Molecular cloning and complete nucleotide sequence of the attenuated rabies vir  
A:Reference number: A34746; MUID:90223994; PMID:2139267  
A:Accession: A34746  
A:Molecule type: genomic RNA  
A:Residues: 1-450 <CON>  
A:Cross-references: GB:M31046; NID:9333556; PIDN:AAA47199.1; PID:9333557  
C:Genetics:  
A:Gene: N  
C:Superfamily: rabies virus nucleoprotein  
C:Keywords: nucleoprotein

Query Match 60.0%; Score 39; DB 1; Length 450;  
Best Local Similarity 63.6%; Pred. No. 36;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 1 VNWGLAACAP 11  
: | | | | | | |  
DB 325 LNATVIAACAP 335

## RESULT 15

VHVNAY  
nucleoprotein - rabies virus (strain AVO1)  
C:Species: rabies virus  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Jul-1999  
C:Accession: S07813  
R:Poch, O.; Torco, N.; Keith, G.  
Biochimie 70, 1019-1029, 1988  
A:Title: Sequence of the 3386 3' nucleotides of the genome of the AVO1 strain rabies  
A:Reference number: S07813; MUID:89150295; PMID:3147658  
A:Accession: S07813  
A:Molecule type: genomic RNA  
A:Residues: 1-450 <POC>  
A:Cross-references: EMBL:X13357; NID:961809; PIDN:CAA31733.1; PID:961810  
C:Genetics:  
A:Gene: N  
C:Superfamily: rabies virus nucleoprotein  
C:Keywords: nucleoprotein

Query Match 60.0%; Score 39; DB 1; Length 450;  
Best Local Similarity 63.6%; Pred. No. 36;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VNWGLAACAP 11  
: | | | | | | |  
DB 325 LNATVIAACAP 335

Search completed: May 20, 2003, 16:37:06  
Job time: 7.28571 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:05:42 ; Search time 3.5 Seconds  
(without alignments)  
154.055 Million cell updates/sec

Title: US-09-869-155-18

Sequence: 1 VNVGVLAACAPSK 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	61.5	659	1 AMIA_STRPN	P18791 streptococ
2	39	60.0	409	1 GUNB_RUMAL	P23661 ruminalococ
3	39	60.0	450	1 NCAP_RABVA	P15197 rabies viru
4	39	60.0	450	1 NCAP_RABVP	P06025 rabies viru
5	39	60.0	450	1 NCAP_RABVP	P16285 rabies viru
6	39	60.0	450	1 NCAP_RABVP	P09110 rabies viru
7	38	60.0	450	1 NCAP_RABVP	P16285 rabies viru
8	38	58.5	150	1 PAL_PASMU	O51886 pasteurella
9	38	58.5	150	1 YN06_MYCTU	P73303 streptocyst
10	37	56.9	147	1 RL15_SYNY3	P27771 streptocyst
11	37	56.9	263	1 PROC_TREPA	P71694 mycobacteri
12	37	56.9	459	1 YE43_MYCTU	P47950 neurospora
13	37	56.9	1117	1 CYT4_NEUCR	P12690 streptomyce
14	36	55.4	154	1 SARA_STRCO	P31468 escherichia
15	36	55.4	155	1 YIE1_ECOLI	O28680 oryctolagus
16	36	55.4	372	1 CD14_RABIT	O50337 mycoplasma
17	35	53.8	217	1 YF90_MYCPN	O69556 mycobacteri
18	35	53.8	517	1 MURF_MYCLE	P48813 saccharomyc
19	35	53.8	663	1 GNP1_YEAST	P28947 equine harp
20	35	53.8	887	1 OBP_HSEVB	P42671 avian adeno
21	35	53.8	942	1 HEX_ADEG1	O06081 mycobacteri
22	35	53.8	955	1 MML3_MYCLE	P46531 homo sapien
23	35	53.8	2556	1 NTCL_HUMAN	P31306 streptococ
24	34	52.3	55	1 SARA_STRCO	O9112 rana catesb
25	34	52.3	173	1 CRAB_RANCA	P40131 salmonella
26	34	52.3	204	1 RK3_GUTTH	O46594 guillardi
27	34	52.3	219	1 FLGA_SALTY	P19641 escherichia
28	34	52.3	323	1 ISPB_ECOLI	O42266 icetulus p
29	34	52.3	346	1 OPRP_ICTPU	P00737 homo sapien
30	34	52.3	347	1 HPT1_HUMAN	P50417 ateles geof
31	34	52.3	347	1 HPT1_HUMAN	P00739 homo sapien
32	34	52.3	348	1 HPT1_HUMAN	Q13351 homo sapien
33	34	52.3	362	1 KLF1_HUMAN	

34	34	52.3	375	1 CD14_HUMAN	P08571 homo sapien
35	34	52.3	403	1 MHPT_ECOLI	P77589 escherichia
36	34	52.3	406	1 HPT2_HUMAN	P00738 homo sapien
37	34	52.3	407	1 MOSC_RHIME	O07609 rhizobium m
38	34	52.3	414	1 FTSW_ECOLI	P16457 escherichia
39	34	52.3	473	1 DCOR_CANAL	P78599 candida alb
40	34	52.3	484	1 SCAT_ECOLI	P39301 escherichia
41	34	52.3	751	1 USM3_DROME	O9yce9 drosophila
42	33	50.8	33	1 TXH1_SELHU	P56676 selencocmi
43	33	50.8	158	1 RM19_YEAST	P53875 saccharomyc
44	33	50.8	176	1 KITH_SFVKA	P07605 shope fibro
45	33	50.8	214	1 PCTT_RAT	P53809 rattus norv

## ALIGNMENTS

RESULT 1	ID	AMIA_STRPN	STANDARD:	PRT:	659 AA.
AC	P18791	AMIA_STRPN	STANDARD:	PRT:	659 AA.
DT	01-NOV-1990	(Rel. 16, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Oligopeptide-binding protein amia precursor.				
GN	AMIA OR SPI891.				
OS	Streptococcus pneumoniae.				
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
OC	Streptococcus.				
OX	NCBI_TaxID=1313;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=R800:				
RX	MEDLINE=90279506; PubMed=2352474;				
RA	Allong G., Trombe M.C., Claverys J.-P.;				
RT	"The ami locus of the Gram-positive bacterium Streptococcus pneumoniae				
RT	is similar to binding protein-dependent transport operons of gram-				
RT	negative bacteria.";				
RL	Mol. Microbiol. 4:633-644(1990).				
RN	[2]				
RP	REVISIONS.				
RA	Claverys J.-P.;				
RL	Submitted (SEP-1992) to the EMBL/Genbank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=TIGR4;				
RX	MEDLINE=21357209; PubMed=11463916;				
RA	Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,				
RA	Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,				
RA	Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,				
RA	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,				
RA	Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,				
RA	McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,				
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,				
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;				
RT	"Complete genome sequence of a virulent isolate of Streptococcus				
RT	pneumoniae.";				
RT	Science 293:498-506(2001).				
RN	[4]				
RP	SEQUENCE OF 1-28 FROM N.A.				
RC	STRAIN=R800:				
RX	MEDLINE=90060772; PubMed=2684766;				
RA	Martin B., Allong G., Boucraut C., Claverys J.-P.;				
RT	"The difficulty of cloning Streptococcus pneumoniae mal and ami loci				
RT	in Escherichia coli: toxicity of malx and amia gene products.";				
RT	Gene 80:227-238(1989).				
RN	[5]				
RP	PARTIAL SEQUENCE.				
RX	MEDLINE=89091147; PubMed=3208757;				
RA	Gilson E., Allong G., Schmidt T., Claverys J.-P., Dudler R.,				
RA	Hofnung M.;				
RT	"Evidence for high affinity binding-protein dependent transport				
RT	systems in Gram-positive bacteria and in Mycoplasma.";				

```
RL EMBL J. 7:3971-3974(1988).
CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC FOR OLIGOPEPTIDES; PROBABLY AN OLIGOPEPTIDE BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC -1- CAUTION: THE REVISED SEQUENCE OF AMIA NOW INCLUDES, IN THE
CC C-TERMINAL SECTION, THE SEQUENCE OF AN ORF WHICH WAS PREVIOUSLY
CC KNOWN AS AMIB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

Query Match: 60.0%; Score 39; DB 1; Length 450;  
Best Local Similarity 63.6%; Pred. No. 11;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNWGVLAACAP 11  
: | : | | | | | |  
DB 325 LNATVIAACAP 335

RESULT 4  
NCAP\_RABVP STANDARD; PRT; 450 AA.  
ID NCAP\_RABVP  
AC P06023;  
DT 01-OCT-1994 (Rel. 30, Created)  
RT 15-OCT-2001 (Rel. 40, Last sequence update)  
DE Nucleocapsid protein (Nucleoprotein).  
GN N.  
OS Rabies virus (strain Ontario fox).  
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
OC Rhabdoviridae; Lyssavirus.  
OX NCBI\_TaxID=37132;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93260396; PubMed=8492088;  
RA Nadin-Davis S.A., Casey G.A., Mandeler A.;  
RT "Identification of regional variants of the rabies virus within the  
Canadian province of Ontario";  
J. Gen. Virol. 74:829-837(1993).  
RN [2]  
RP REVISIONS TO 444 AND 450.  
RA Nadin-Davis S.A.;  
RT Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L20673; AAA03482.2; -  
DR EMBL: L20675; AAA03484.2; -  
DR EMBL: L20676; AAA03485.2; -  
DR EMBL: L20672; AAA82762.1; -  
DR InterPro: IPR000448; Rhabd\_nucleocap.  
DR Pfam: PF00945; Rhabd\_nucleocap; 1.  
DR ProDom: PD002087; Rhabd\_nucleocap; 1.  
KW Nucleocapsid.  
SQ SEQUENCE 450 AA; 50593 MW; D224ACADB773B19 CRC64;

Query Match: 60.0%; Score 39; DB 1; Length 450;  
Best Local Similarity 63.6%; Pred. No. 11;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNWGVLAACAP 11  
: | : | | | | | |  
DB 325 LNATVIAACAP 335

RESULT 5  
NCAP\_RABVP STANDARD; PRT; 450 AA.  
ID NCAP\_RABVP  
AC P06023;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Nucleocapsid protein (Nucleoprotein).  
GN N.  
OS Rabies virus (strain Pasteur / PV).  
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
OC Rhabdoviridae; Lyssavirus.

OX NCBI\_TaxID=103929;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86176757; PubMed=3008096;  
RA Tordo N., Poch O., Ermine A., Keith G.;  
RT "Primary structure of leader RNA and nucleoprotein genes of the  
RT rabies genome: segmented homology with VSV";  
Nucleic Acids Res. 14:2671-2683(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=8623343; PubMed=3459163;  
RA Tordo N., Poch O., Ermine A., Keith G., Rougeon F.;  
RT "Walking along the rabies genome: is the large G-L intergenic region  
RT a remnant gene?";  
Proc. Natl. Acad. Sci. U.S.A. 83:3914-3918(1986).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X03673; CAA27308.1; -  
DR EMBL: M13215; AAA47215.1; -  
DR EMBL: A14671; CAA01189.1; -  
DR EMBL: A14407; CAA01155.1; -  
DR PIR: D26275; VHNVRV.  
DR InterPro: IPR000448; Rhabd\_nucleocap.  
DR Pfam: PF00945; Rhabd\_nucleocap; 1.  
DR ProDom: PD002087; Rhabd\_nucleocap; 1.  
KW Nucleocapsid.  
SQ SEQUENCE 450 AA; 50605 MW; 670F9EC48985C3BB CRC64;

Query Match: 60.0%; Score 39; DB 1; Length 450;  
Best Local Similarity 63.6%; Pred. No. 11;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNWGVLAACAP 11  
: | : | | | | | |  
DB 325 LNATVIAACAP 335

RESULT 6  
NCAP\_RABVS STANDARD; PRT; 450 AA.  
ID NCAP\_RABVS  
AC P16285;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-AUG-1990 (Rel. 15, Last annotation update)  
DE Nucleocapsid protein (Nucleoprotein).  
GN N.  
OS Rabies virus (strain SAD B19).  
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
OC Rhabdoviridae; Lyssavirus.  
OX NCBI\_TaxID=11300;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90223994; PubMed=2139267;  
RA Conzelmann K.-K., Cox J.H., Schneider L.G., Thiel H.-J.;  
RT "Molecular cloning and complete nucleotide sequence of the attenuated  
RT rabies virus SAD B19";  
Virology 175:485-499(1990).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M31046; AAA47199.1; -  
 DR PIR: A34746; VHVNSB.  
 DR InterPro: IPR000448; Rhabd\_nucleocap.  
 DR Pfam: PF00945; Rhabd\_nucleocap; 1.  
 DR ProDom: PD002087; Rhabd\_nucleocap; 1.  
 KW Nucleocapsid.  
 SQ SEQUENCE 450 AA; 50603 MW; 783BF3E01E7BE325 CRC64;

Query Match 60.0%; Score 39; DB 1; Length 450;  
 Best Local Similarity 63.6%; Pred. No. 11;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VNVGLAACAP 11  
 : 1:|||||  
 Db 325 LNATVIAACAP 335

## RESULT 7

NCAP\_RABVU STANDARD: PRT; 450 AA.

AC Q09110;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DE 01-FEB-1995 (Rel. 31, Last annotation update)

DE Nucleocapsid protein (Nucleoprotein).

GN N.

OS Rabies virus (strain Ontario skunk).

OC Viruses: ssRNA negative-strand viruses; Mononegavirales;

OC Rhabdoviridae; Lyssavirus.

OX NCBI\_TaxID=39005;

RN [..]

RP SEQUENCE FROM N.A.

RA MEDLINE-93260396; PubMed-8492088;

RA Nadin-Davis S.A., Casey G.A., Wandeler A.;

RT "Identification of regional variants of the rabies virus within the

RT Canadian province of Ontario";

RL J. Gen. Virol. 74:829-837(1993).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: L20671; AAA47220.1; -

DR InterPro: IPR000448; Rhabd\_nucleocap.

DR Pfam: PF00945; Rhabd\_nucleocap; 1.

DR ProDom: PD002087; Rhabd\_nucleocap; 1.

KW Nucleocapsid.

SQ SEQUENCE 450 AA; 50611 MW; D0FAC9A9A7773C69 CRC64;

Query Match 60.0%; Score 39; DB 1; Length 450;  
 Best Local Similarity 63.6%; Pred. No. 11;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VNVGLAACAP 11  
 : 1:|||||  
 Db 325 LNATVIAACAP 335

## RESULT 8

PAL\_PASMU STANDARD: PRT; 150 AA.

AC Q51886;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane protein P6 precursor (OMP P6) (P6-like) (Peptidoglycan-

DE associated lipoprotein).

DE 3AL OR PM0966.

OS Pasteurella multocida.

OC Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Pasteurella.

OX NCBI\_TaxID=747;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-T16;

RX MEDLINE-95172751; PubMed-7868272;

RA Kasten R.W., Hansen L.M., Hinojoza J., Bleher D., Ruehl W.W.,

RT Hirsch D.C.;

RT "Pasteurella multocida produces a protein with homology to the P6

RT outer membrane protein of Haemophilus influenzae";

RL Infect. Immun. 63:989-993(1995).

RN (2)

RP SEQUENCE FROM N.A.

RC STRAIN-Pm70;

RX MEDLINE-21145866; PubMed-11248100;

RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida Pm70";

RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID

CC ANCHOR (By similarity).

CC -1- SIMILARITY: TO OTHER PAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: U16849; AAA79373.1; -

DR EMBL: AE006136; AK03050.1; -

DR InterPro: IPR001145; Bac\_OmpA.

DR Pfam: PF00661; OmpA; 1.

DR PRINTS: PR01021; OMPADOMAIN.

DR ProDom: PD000930; Bac\_OmpA; 1.

DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.

DR PROSITE: PS01068; OMPA; 1.

KW Outer membrane: Signal; Lipoprotein: Complete proteome.

FT SIGNAL 1 19 BY SIMILARITY.

FT CHAIN 20 150 OUTER MEMBRANE PROTEIN P6.

FT LIPID 20 20 N-ACYL DIGLUCERIDE (BY SIMILARITY).

FT DOMAIN 82 126 OMPA-LIKE.

SQ SEQUENCE 150 AA; 16213 MW; 170E7B13D2E9ED6C CRC64;

Query Match 58.5%; Score 38; DB 1; Length 150;  
 Best Local Similarity 66.7%; Pred. No. 6; Gaps 0;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 NNGVLAACAPSK 13  
 : 1:|||||  
 Db 13 SVAVLAACGSSK 24

## RESULT 9

YN06\_MYCTU STANDARD: PRT; 208 AA.

AC Q50659;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical protein RV2306c precursor.

GN RV2306C OR MT2363 OR MTCY339.03.

OS Mycobacterium tuberculosis.

OC Bacteria: Actinobacteria; Actinobacteria (class); Actinobacteridae;

OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1773;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-H37Rv;

RX MEDLINE-98293987; PubMed-9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

```

RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Razaandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrett B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RT Nature 393:537-544(1998).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=CDC 1591 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.,
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; 277163; CAB00992.1; -
CC DR EMBL; AE007078; AAK46649.1; -
CC DR TIGR; MT2363; -
CC DR TubercuList; RV2306c; -
CC DR Hypothetical protein; Signal; Complete proteome.
CC KW SIGNAL
CC FT CHAIN 17 208
CC FT SEQUENCE 208 AA; 22521 MW; DE913A5D2C34DA2F CRC64;
SQ
Query Match 58.5%; Score 38; DB 1; Length 208;
Best Local Similarity 54.5%; Pred. No. 8.8;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 3 VGVLAACAPSK 13
DB 131 LGVTPACAPAE 141
RESULT 10
RL5_SYNY3
ID RL5_SYNY3 STANDARD; PRT; 147 AA.
AC P73303;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE 50S ribosomal protein L15.
GN RPL0 OR RPL15 OR SL1813.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugitani M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RT DNA Res. 3:109-136(1996).
CC -I- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA

```

```

CC (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; D90905; BAA17332.1; -
CC DR InterPro; IPR001196; Ribosomal_L15.
CC DR Pfam; PF00256; L15; 1.
CC DR Pfam; PF01305; Ribosomal_L15; 1.
CC DR TIGRPFAM; TIGR01071; rpl0_bact; 1.
CC DR PROSITE; PS00475; RIBOSOMAL_L15; 1.
CC KW Ribosomal protein; rRNA-binding; Complete proteome.
CC SEQUENCE 147 AA; 15195 MW; 18DC0B2F69B1AE09 CRC64;
SQ
Query Match 56.9%; Score 37; DB 1; Length 147;
Best Local Similarity 66.7%; Pred. No. 9.8;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 VNVGVLAACAPS 12
DB 78 VNVGKLAGLAPN 89
RESULT 11
PROC_TREPA
ID PROC_TREPA STANDARD; PRT; 263 AA.
AC P27771; O83775;
DT 01-AUG-1992 (rel. 23, Created)
DT 15-DEC-1998 (rel. 37, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
GN PROC OR TP0797.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=90264287; PubMed=2188947;
RA Cheradimi F.C., Hobbs M.M., Stamm L.V., Bassford P.J. Jr.;
RT "Complementation of an Escherichia coli proc mutation by a gene cloned
RT from Treponema pallidum."
RT J. Bacteriol. 172:2996-3002(1990).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uitterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RT Science 281:375-388(1998).
RL -I- CATALYTIC ACTIVITY: L-proline + MAD(P)(+) = 1-pyrolone-5-
CC carboxylate + NAD(P)H.
CC -I- PATHWAY: Proline biosynthesis; third (last) step.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: BELONGS TO THE PYROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M73825; AAA2478.1; -  
DR EMBL: AE001250; AAC65760.1; -  
DR PIR: S27734; S27734.  
DR TIGR: TP0797; -  
DR InterPro: IPR000304; P5CR.  
DR Pfam: PF01089; P5CR.1.  
DR PROSITE: PS00521; P5CR.1.  
DR Oxidoreductase; Proline biosynthesis; NADP: Complete proteome.  
FT CONFLICT 1 65  
FT MVTGIFGAGRALAGLVAAGALQAQVYACALNDKLR  
FT AACTSISIGACASYQELVOKSEMI -> MTMDEVLQMDG  
FT SGTGTEKNGVF (IN REF. 1)  
FT VLRDRSFQCKVLISLA -> GTRASPIQDSALSC (IN  
FT CONFLICT 77 93  
FT REF. 1)  
FT VLRDRSFQCKVLISLA -> GTRASPIQDSALSC (IN  
FT CONFLICT 256 260  
FT VLRAL -> CRWLS (IN REF. 1).  
FT SEQUENCE 263 AA; 27645 MW; ED4AD7C54BAF9D61 CRC64;

Query Match 56.9%; Score 37; DB 1; Length 263;  
Best Local Similarity 66.7%; Pred. NO. 16;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 4 GVLACAPS 12  
Db 169 GVLACAPA 177

RESULT 12  
YE43\_MYCTU STANDARD; PRT; 459 AA.  
ID YE43\_MYCTU  
AC P71694; 006833;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 13-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein RV1425.  
GN RV1425 OR MT1468 OR MTCY21B4.43 OR MTCY493.29C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinobacteriales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
RA "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains.";  
RT Submitted (APR-2001) to the EMBL/GenBank/DDRF databases.  
CC -1- SIMILARITY: BELONGS TO THE UPF0089 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: 295844; CAB09245.1; -  
DR EMBL: AE007017; AAK45733.1; ALT\_INIT.  
DR TIGR: MT1468; -  
DR Tuberculin; Rv1425; -  
DR InterPro: IPR004235; UPF0089.  
DR Pfam: PF03007; UPF0089.1.  
DR Hypothetical protein; Complete proteome.  
SO SEQUENCE 459 AA; 50062 MW; F436D6B588562263 CRC64;

Query Match 56.9%; Score 37; DB 1; Length 459;  
Best Local Similarity 58.3%; Pred. NO. 27;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 VNVGVLACAPS 12  
Db 21 MHVGLALICDPS 32

RESULT 13  
CYT4\_NEUCR STANDARD; PRT; 1117 AA.  
ID CYT4\_NEUCR  
AC P47950;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Mitochondrial protein cyt-4.  
GN CYT-4.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A;  
RX MEDLINE=92179246; PubMed=1311848;  
RA Turc B., Dobinson K.F., Serizawa N., Lambowitz A.M.;  
RT "A protein required for RNA processing and splicing in Neurospora  
RT mitochondria is related to gene products involved in cell cycle  
RT protein phosphatase functions.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:1676-1680(1992).  
CC -1- FUNCTION: REQUIRED FOR RNA 5' AND 3' END PROCESSING AND SPLICING.  
CC MAY ACT ON THE RNA PROCESSING ENZYME DIRECTLY, OR IT MAY ACT ON  
CC OTHER REGULATORY MOLECULES, WHICH INFLUENCE THE ACTIVITY OR  
CC SYNTHESIS OF THESE ENZYMES.  
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M80735; AAA33608.1; -  
DR PIR: A38227; A38227.  
DR InterPro: IPR001900; Ribonuclease\_II.  
DR Pfam: PF00773; RNB.1.  
DR PROSITE: PS01175; RIBONUCLEASE\_II; 1.  
KM Mitochondrion; mRNA processing.  
SO SEQUENCE 1117 AA; 125103 MW; 93D1832D9128B20B CRC64;

Query Match 56.9%; Score 37; DB 1; Length 1117;  
Best Local Similarity 63.6%; Pred. NO. 59;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 NCVLAACAPS 12  
1:|||||1|  
Db 204 NCVLAACAPS 214

RESULT 14  
SAPA\_STRCO STANDARD; PRT; 154 AA.  
AC P12690; Q9RJS7; 01-OCT-1989 (Rel. 12, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Spore-associated protein A precursor.  
GN SAPA OR SC00409 OR SCF51.08C.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Crolin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Ruter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
RN [2]  
RP SEQUENCE OF 1-116 FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=88169521; PubMed=2450872;  
RA Guirarro J., Santamaria R., Schauer A., Losick R.;  
RT "Promoter determining the timing and spatial localization of  
transcription of a cloned Streptomyces coelicolor gene encoding a  
spore-associated polypeptide.";  
RT J. Bacteriol. 170:1895-1901(1988).  
CC -|- SUBCELLULAR LOCATION: SPORE SURFACE.  
CC -|- DEVELOPMENTAL STAGE: TRANSCRIPTION OF THIS PROTEIN IS INDUCED AT  
THE TIME OF APPEARANCE OF AERIAL MYCELIIUM BEFORE THE BEGINNING OF  
SPORE FORMATION.  
CC -|- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 79  
ONWARD DUE TO A FRAMESHIFT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; A1132707; CAB59706.1; -  
DR EMBL; M20145; AAA26812.1; ALT\_FRAME.  
KW Sporulation; Signal; Complete proteome.  
FT SIGNAL 27  
FT CHAIN 1 154 SPORE-ASSOCIATED PROTEIN A.  
SQ SEQUENCE 154 AA; 15641 MW; B0B72EAF54361547 CRC64;

Query Match 55.4%; Score 36; DB 1; Length 154;  
Best Local Similarity 66.7%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 NCVLAACAPS 13  
1:|||||1|

Db 136 NCGALAAKAPK 147

RESULT 15  
YIEI\_ECOLI STANDARD; PRT; 155 AA.  
AC P31468; 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yieI.  
GN YIEI OR B3716.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=93315143; PubMed=7686882;  
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;  
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli  
genome: organizational symmetry around the origin of replication.";  
RL Genomics 16:551-561(1993).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L10328; AAA62067.1; -  
DR EMBL; AE000448; AAC76739.1; -  
DR Ecogene; Egl1726; yieI.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 155 AA; 16683 MW; 779FDA7162CF501D CRC64;

Query Match 55.4%; Score 36; DB 1; Length 155;  
Best Local Similarity 60.0%; Pred. No. 16;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 GYLAACAPS 13  
1:|||||1|  
Db 63 GYLAACAPS 72

Search completed: May 20, 2003, 16:35:46  
Job time : 5.5 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:19:09 ; Search time 16.0714 Seconds  
(without alignments)  
166.669 Million cell updates/sec

Title: US-09-869-155-18  
Perfect score: 65  
Sequence: 1 VNVGYLACAPSK 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	63.1	169	16 Q9RV50	Q9RV50 deinoococcus
2	41	63.1	419	5 Q9NA38	Q9NA38 caenorhabdi
3	41	63.1	420	5 Q9NA61	Q9NA61 caenorhabdi
4	41	63.1	450	12 Q8OY66	Q8OY66 rabies viru
5	40	61.5	228	16 Q9RUB6	Q9RUB6 deinoococcus
6	40	61.5	349	16 Q931D0	Q931D0 rhizobium m
7	40	61.5	637	5 Q9GZ30	Q9GZ30 trypsinosoma
8	39	60.0	257	16 Q8YMX1	Q8YMX1 anabaena sp
9	39	60.0	288	5 Q45711	Q45711 caenorhabdi
10	39	60.0	290	10 Q9SVB1	Q9SVB1 arabidopsis
11	39	60.0	341	10 Q9A456	Q9A456 rabies viru
12	39	60.0	445	12 Q99106	Q99106 rabies viru
13	39	60.0	450	12 Q85412	Q85412 rabies viru
14	39	60.0	450	12 Q55611	Q55611 rabies viru
15	39	60.0	450	12 Q9W9S4	Q9W9S4 rabies viru
16	39	60.0	450	12 Q39473	Q39473 rabies viru

17	39	60.0	450	12 Q86407	Q86407 rabies viru
18	39	60.0	450	12 Q86408	Q86408 rabies viru
19	39	60.0	450	12 Q86409	Q86409 rabies viru
20	39	60.0	450	12 Q86410	Q86410 rabies viru
21	39	60.0	450	12 Q86413	Q86413 rabies viru
22	39	60.0	450	12 Q86414	Q86414 rabies viru
23	39	60.0	450	12 Q86415	Q86415 rabies viru
24	39	60.0	450	12 Q86416	Q86416 rabies viru
25	39	60.0	450	12 Q86417	Q86417 rabies viru
26	39	60.0	450	12 Q86418	Q86418 rabies viru
27	39	60.0	450	12 Q86420	Q86420 rabies viru
28	39	60.0	450	12 Q86421	Q86421 rabies viru
29	39	60.0	450	12 Q86422	Q86422 rabies viru
30	39	60.0	450	12 Q86423	Q86423 rabies viru
31	39	60.0	450	12 Q86424	Q86424 rabies viru
32	39	60.0	450	12 Q86425	Q86425 rabies viru
33	39	60.0	450	12 Q86426	Q86426 rabies viru
34	39	60.0	450	12 Q86427	Q86427 rabies viru
35	39	60.0	450	12 Q86429	Q86429 rabies viru
36	39	60.0	450	12 Q86430	Q86430 rabies viru
37	39	60.0	450	12 Q86431	Q86431 rabies viru
38	39	60.0	450	12 Q86432	Q86432 rabies viru
39	39	60.0	450	12 Q86433	Q86433 rabies viru
40	39	60.0	450	12 Q86434	Q86434 rabies viru
41	39	60.0	450	12 Q86435	Q86435 rabies viru
42	39	60.0	450	12 Q86436	Q86436 rabies viru
43	39	60.0	450	12 Q86437	Q86437 rabies viru
44	39	60.0	450	12 Q86438	Q86438 rabies viru
45	39	60.0	450	12 Q86441	Q86441 rabies viru

## ALIGNMENTS

### RESULT 1

Q9RV50 ID Q9RV50 PRELIMINARY; PRT; 169 AA.  
AC Q9RV50;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein DR1180.  
GN DR1180.  
OS Deinoococcus radiodurans.  
OC Bacteria; Thermus/Deinoococcus group; Deinoococi; Deinoococcales;  
OC Deinoococcaceae; Deinoococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SRRATN-RI;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Olin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "genome sequence of the radioresistant bacterium Deinoococcus  
RT radiodurans RI.";  
RL Science 286:1571-1577(1999).  
DR EMBL; AF001967; AAF10759.1; -  
DR TIGR; DR1180; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 169 AA; 17313 MW; 22D92D6C79C5A0C5 CRC64;

Query Match 63.1%; Score 41; DB 16; Length 169;

Best Local Similarity 70.0%; Pred. No. 8.8;

Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 VGVLAACAPS 12  
11:11:11:  
Db 13 VGLIASCAPA 22

```
RESULT 2
Q9N38      PRELIMINARY;      PRT;      419 AA.
AC  Q9N38;
DT  01-OCT-2000 (TREMBlrel. 15, Created)
DT  01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE  01-JUN-2001 (TREMBlrel. 17, Last annotation update)
GN  Y3F8A.9 protein.
OS  Caenorhabditis elegans.
OC  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC  Rhabditidae; Pelodierinae; Caenorhabditis.
OX  NCBI_TaxID=6239;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Matthews L.;
RL  Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RX  MEDLINE=99069613; PubMed=9851916;
RT  [2]
"Genome sequence of the nematode C. elegans: A platform for
investigating biology."
RL  Science 282:2012-2018(1998).
DR  EMBL; AL132862; CAB70227.1; -.
DR  InterPro; IPR003341; DUF139.
DR  Pfam; PF02363; DUF139; 11.
SQ  SEQUENCE 419 AA; 45397 MW; F62317E9E7B408DF CRC64;

Query Match      63.1%; Score 41; DB 5; Length 419;
Best Local Similarity 63.6%; Pred. No. 21;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY  1 VNVGLAACAP 11
Db  194 VNLGISAQCAP 204

RESULT 3
Q9N61      PRELIMINARY;      PRT;      420 AA.
AC  Q9N61;
DT  01-OCT-2000 (TREMBlrel. 15, Created)
DT  01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE  01-DSC-2001 (TREMBlrel. 19, Last annotation update)
GN  Y3F8A.8 protein.
OS  Caenorhabditis elegans.
OC  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC  Rhabditidae; Pelodierinae; Caenorhabditis.
OX  NCBI_TaxID=6239;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Matthews L.;
RL  Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RX  MEDLINE=99069613; PubMed=9851916;
RT  [2]
"Genome sequence of the nematode C. elegans: A platform for
investigating biology."
RL  Science 282:2012-2018(1998).
DR  EMBL; AL132862; CAB60533.1; -.
DR  InterPro; IPR003341; DUF139.
DR  Pfam; PF02363; DUF139; 11.
SQ  SEQUENCE 420 AA; 45412 MW; 51BAB4ED5EA9E538 CRC64;

Query Match      63.1%; Score 41; DB 5; Length 420;
Best Local Similarity 63.6%; Pred. No. 21;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
QY  1 VNVGLAACAP 11
Db  195 VNLGISAQCAP 205

RESULT 4
Q80Y66     PRELIMINARY;      PRT;      450 AA.
AC  Q80Y66;
DT  01-JUN-2002 (TREMBlrel. 21, Created)
DT  01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE  01-JUN-2002 (TREMBlrel. 21, Last annotation update)
GN  Nucleoprotein.
OS  Rabies virus.
OC  Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC  Rhabdoviridae; Lyssavirus.
OX  NCBI_TaxID=11292;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Jacobs J.A., Theron J., Nel L.H.;
RT  "In vitro RNA- and protein-binding characteristics of the
nucleoprotein of a South African viverrid rabies virus isolate."
RL  Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF467949; AAL76066.1; -.
SQ  SEQUENCE 450 AA; 50606 MW; E834525A67FC7D74 CRC64;

Query Match      63.1%; Score 41; DB 12; Length 450;
Best Local Similarity 63.6%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY  1 VNVGLAACAP 11
Db  325 LMAVIAACAP 335

RESULT 5
Q9RUB6     PRELIMINARY;      PRT;      228 AA.
AC  Q9RUB6;
DT  01-MAY-2000 (TREMBlrel. 13, Created)
DT  01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE  01-MAR-2002 (TREMBlrel. 20, Last annotation update)
GN  Hypothetical protein DR1474.
OS  Deinococcus radiodurans.
OC  Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC  Deinococcaceae; Deinococcus.
OX  NCBI_TaxID=1299;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  SPRAIN R.;
RX  MEDLINE=20036896; PubMed=10567266;
RA  White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA  Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA  Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA  Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA  Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA  Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA  Fraser C.M.;
RT  "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1."
RL  Science 286:1571-1577(1999).
DR  EMBL; AE001992; AAF11048.1; -.
DR  TIGR; DR1474; -.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 228 AA; 24659 MW; D26BA136F57A4087 CRC64;

Query Match      61.5%; Score 40; DB 16; Length 228;
Best Local Similarity 77.8%; Pred. No. 18;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY  3 VGVIAACAP 11
```

Db 14 VGLLASCAP 22

## RESULT 6

ID 0931D0 PRELIMINARY; PRT; 349 AA.

AC 0931D0; 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Selenide, water dikinase (selenophosphate synthetase), Seld (EC 2.7.9.3).  
 GN Seld OR RA0014 OR SMA0028.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Plasmid pSymba (megaplasmid 1).  
 OC Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 NCBI\_TaxID=382;  
 RX SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RA MEDLINE=21396509; PubMed=11481432;  
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
 RA Bariloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,  
 RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,  
 RA Kallman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,  
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;  
 RT "Nucleotide sequence and predicted functions of the entire  
 RT Sinorhizobium meliloti pSymba megaplasmid.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).  
 DR EMBL: AE007196; AAK6472.1; -;  
 DR InterPro: IPR000728; AIRS.related.  
 DR InterPro: IPR004536; Seld.  
 DR Pfam: PF02769; AIRS.C.1.  
 DR TIGRfams: TIGR00476; Seld; 1.  
 DR Kinase; Transferase; Plasmid; Complete proteome.  
 SQ SEQUENCE 349 AA; 35833 MW; 1196A5D9D6F87CF3 CRC64;

Query Match Best Local Similarity 61.5%; Score 40; DB 16; Length 349;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVLACAPSK 13  
 Db 307 GLVACAPSK 316

## RESULT 7

ID 09GZ30 PRELIMINARY; PRT; 637 AA.

AC 09GZ30; 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Hypothetical 72.2 kDa protein.  
 OS Trypanosoma cruzi.  
 OC Eukaryota, Euzlenozoa, Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 NCBI\_TaxID=5693;  
 RN NCB1\_TaxID=5693;  
 RP SEQUENCE FROM N.A.  
 RA Andersson B.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF242860; AAF98148.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 637 AA; 72214 MW; 6470347EF695D33F CRC64;

Query Match Best Local Similarity 61.5%; Score 40; DB 5; Length 637;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 NCVLAACAPS 12  
 I:|||||:

Db 473 NTNVAACAPT 483

## RESULT 8

ID 08YMX1 PRELIMINARY; PRT; 257 AA.

AC 08YMX1; 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.  
 GN ALR4803.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria: Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 NCBI\_TaxID=103690;  
 RX SEQUENCE FROM N.A.  
 RP MEDLINE=21595285; PubMed=11759840;  
 RA Kaneo T., Nakamura Y., Wolk C.P., Kawasaki K., Kimura T.,  
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawasaki K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL: AP003597; BAB76502.1; -;  
 DR InterPro: IPR002529; FAA\_hydrolase.  
 DR Pfam: PF01557; FAA\_hydrolase; 1.  
 KW Isomerase; Complete proteome.  
 SQ SEQUENCE 257 AA; 28446 MW; 7B9F7470042323A5 CRC64;

Query Match Best Local Similarity 60.0%; Score 39; DB 16; Length 257;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 VLAACAPSK 13  
 Db 51 ILAPCAPSK 59

## RESULT 9

ID 045711 PRELIMINARY; PRT; 288 AA.

AC 045711; 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE R09B3.1 protein.  
 GN R09B3.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodermidae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 RN NCB1\_TaxID=6239;  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smalton N., Smith A., Sonhammer E., Straden R., Sulston J.,  
 RA Telford-Mleg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";

```

RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Tremblay S., Masson J.-Y., Ramotar D.;
RT "The exonuclease III family of DNA repair enzymes has a new homolog in
RL CeeXIII from Caenorhabditis elegans."
DR EMBL: Z81108; CAB03235.1; -
DR EMBL: AF034258; AAC82328.1; -
DR HSSP: P27695; 1E9N.
DR InterPro: IPR004442; Exonuclease_III.
DR InterPro: IPR004808; ExoIII_xth.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR TIGRfams: TIGR00195; exonuclease_III; 1.
DR TIGRfams: TIGR00633; xth; 1.
KM Exonuclease.
SQ SEQUENCE 288 AA; 32898 MW; 104C97DE2B7F9661 CRC64;

Query Match
Best Local Similarity 60.0%; Score 39; DB 5; Length 288;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VGVLAACAPSK 13
   ||| |||
DB 102 VGLSKCAPMK 112

RESULT 10
Q9SVB1 ID Q9SVB1 PRELIMINARY; PRT; 290 AA.
AC Q9SVB1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Hypothetical 33.4 kDa protein.
F23K16.100 OR A74G39470.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL078620; CAB44682.1; -
DR EMBL: AL161595; CAB80610.1; -
DR InterPro: IPR001440; TPR.
KM Hypothetical protein.
SQ SEQUENCE 290 AA; 33444 MW; A76771155D9E72A5 CRC64;

Query Match
Best Local Similarity 60.0%; Score 39; DB 10; Length 290;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 NVGVLAACAPSK 13
   ::||| ||| |||
DB 3 SIGVTALCSPAK 14

```

```

RESULT 11
Q94A56 ID Q94A56 PRELIMINARY; PRT; 341 AA.
AC Q94A56;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE A74C39470/F23K16.100.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamita A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY050356; MAK91373.1; -
DR InterPro: IPR004440; TPR.
DR SMART: SM00028; TPR; 1.
SQ SEQUENCE 341 AA; 39155 MW; 9708D7021C06B5E1 CRC64;

Query Match
Best Local Similarity 60.0%; Score 39; DB 10; Length 341;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 NVGVLAACAPSK 13
   ::||| ||| |||
DB 3 SIGVTALCSPAK 14

```

```

RESULT 12
Q99106 ID Q99106 PRELIMINARY; PRT; 445 AA.
AC Q99106;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Nucleocapsid protein (Nucleoprotein) (Fragment).
OS Rabies virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Rhabdoviridae; Lyssavirus.
OX NCBI_TaxID=11292;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91306453; PubMed=1840709;
RA Sumner J.W., Pekadu M., Shaddock J.H., Esposito J.J., Bellini W.;
RT "Protection of mice with vaccinia virus recombinants that express the
RT rabies nucleoprotein."
RL Virology 185:703-710(1991).
DR EMBL: M61047; AAA47222.1; -
DR InterPro: IPR000448; Rhabd_nucleocap.
DR Pfam: PF00945; Rhabd_nucleocap; 1.
DR Prodom: PD002087; Rhabd_nucleocap; 1.
KM Nucleocapsid.
FT NON_TER
SQ SEQUENCE 445 AA; 50183 MW; 76734579132E3F00 CRC64;

Query Match
Best Local Similarity 60.0%; Score 39; DB 12; Length 445;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNVGVLAACAPSK 11
   :| :| |||||

```

Db 320 LNATVIAACAP 330

RESULT 13

ID 085412 PRELIMINARY; PRT; 450 AA.

AC 085412;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE NUCLEOPROTEIN.

GN N.

OS Rabies virus.

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;

OC Rhabdoviridae; Lyssavirus.

OX NCBI\_TaxID=11292;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-RC-HL;

RX MEDLINE=94353646; PubMed=8073640;

RA Goto H., Minamoto N., Ito H., Sugiyama M., Kinjo T., Mannen K.,

RA Mitune K., Kawato A.,

RT "Nucleotide sequence of the nucleoprotein gene of the RC-HL strain of

RT rabies virus, a seed strain used for animal vaccine production in

RT Japan."

RT Virus genes 8:91-97(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-RC-HL;

RA Minamoto N.;

RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-RC-HL;

RA Minamoto N.;

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: D16331; BAA03838.1; -

DR EMBL: AB009663; BAA24083.1; -

DR InterPro: IPR000448; Rhabd\_nucleocap.

DR Pfam: PF00945; Rhabd\_nucleocap; 1.

DR ProDom: PD002087; Rhabd\_nucleocap; 1.

KW Nucleoprotein.

SQ SEQUENCE 450 AA; 50658 MW; A73F8EB623D8A5A9 CRC64;

Query Match 60.0%; Score 39; DB 12; Length 450;

Best Local Similarity 63.6%; Pred. No. 53;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNVGVLAACAP 11

Db 325 LNATVIAACAP 335

RESULT 14

ID 055611 PRELIMINARY; PRT; 450 AA.

AC 055611;

DT 01-JUN-1998 (TREMBLrel. 06, Created)

DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE NUCLEOPROTEIN.

GN N.

OS Rabies virus.

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;

OC Rhabdoviridae; Lyssavirus.

OX NCBI\_TaxID=11292;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-NISHIGAHARA;

RA Minamoto N.;

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-NISHIGAHARA;

RA Ito N., Kakemizu M., Ito K.A., Yamamoto A., Yoshida Y., Sugiyama M.,

RA Minamoto N.;

RT "A comparison of complete genome sequences of the attenuated RC-HL

RT strain of rabies virus used for production of animal vaccine in Japan,

RT and the parental Nishigahara strain."

RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB010494; BAA24618.1; -

DR EMBL: AB044824; BAA96802.1; -

DR InterPro: IPR000448; Rhabd\_nucleocap.

DR Pfam: PF00945; Rhabd\_nucleocap; 1.

DR ProDom: PD002087; Rhabd\_nucleocap; 1.

KW Nucleoprotein.

SQ SEQUENCE 450 AA; 50720 MW; 7DAFB96AA7055A2C CRC64;

Query Match 60.0%; Score 39; DB 12; Length 450;

Best Local Similarity 63.6%; Pred. No. 53;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNVGVLAACAP 11

Db 325 LNATVIAACAP 335

RESULT 15

ID 09W9S4 PRELIMINARY; PRT; 450 AA.

AC 09W9S4;

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)

DE NUCLEOPROTEIN.

OS Rabies virus.

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;

OC Rhabdoviridae; Lyssavirus.

OX NCBI\_TaxID=11292;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-8661FRA; AND 9223FRA;

RA Bouthy H., Kissi B., Kulonen K., Tordo N., Audry L., Stohr K.;

RT "Evolution of the nucleoprotein gene and host range of old world

RT rabies virus."

RT Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL: U43434; AAD10458.1; -

DR EMBL: U43433; AAD10457.1; -

DR InterPro: IPR000448; Rhabd\_nucleocap.

DR Pfam: PF00945; Rhabd\_nucleocap; 1.

DR ProDom: PD002087; Rhabd\_nucleocap; 1.

KW Nucleocapsid.

SQ SEQUENCE 450 AA; 50662 MW; DCAC028FFBDAF689 CRC64;

Query Match 60.0%; Score 39; DB 12; Length 450;

Best Local Similarity 63.6%; Pred. No. 53;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNVGVLAACAP 11

Db 325 LNATVIAACAP 335

Search completed: May 20, 2003, 16:43:20

Job time : 18.0714 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:35:02 : Search time 6 Seconds  
(without alignments)  
63.750 Million cell updates/sec

Title: US-09-869-155-18  
Perfect score: 65  
Sequence: 1 VNWGVLAACAPSK 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued\_Patents\_AA:\*  
2: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfillsl.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	63.1	25	US-08-139-609-5	Sequence 5, Appl
2	40	61.5	642	US-08-245-511-48	Sequence 48, Appl
3	40	61.5	642	US-08-600-993A-48	Sequence 48, Appl
4	39	60.0	25	US-08-139-609-3	Sequence 3, Appl
5	38	58.5	26	US-08-139-609-12	Sequence 12, Appl
6	36	55.4	1422	US-08-469-260A-83	Sequence 83, Appl
7	36	55.4	2864	US-08-469-260A-394	Sequence 394, Appl
8	35	53.8	52	US-09-461-697-95	Sequence 95, Appl
9	35	53.8	485	US-09-516-914-1	Sequence 1, Appl
10	35	53.8	569	US-08-686-968C-227	Sequence 227, Appl
11	35	53.8	749	US-09-562-737-99	Sequence 99, Appl
12	35	53.8	942	US-09-171-461-12	Sequence 12, Appl
13	35	53.8	2556	US-08-185-432-17	Sequence 17, Appl
14	35	53.8	2556	US-08-083-590A-20	Sequence 20, Appl
15	35	53.8	2556	US-08-532-384-20	Sequence 20, Appl
16	35	53.8	2556	US-08-899-232-2	Sequence 2, Appl
17	34	52.3	334	US-08-484-397A-8	Sequence 8, Appl
18	34	52.3	348	US-08-366-953A-45	Sequence 45, Appl
19	34	52.3	348	US-08-484-397A-2	Sequence 2, Appl
20	34	52.3	348	US-08-484-397A-3	Sequence 3, Appl
21	34	52.3	348	US-08-484-397A-4	Sequence 4, Appl
22	34	52.3	348	US-08-484-397A-5	Sequence 5, Appl
23	34	52.3	348	US-08-484-397A-6	Sequence 6, Appl
24	34	52.3	348	US-08-484-397A-7	Sequence 7, Appl
25	34	52.3	348	US-08-484-397A-27	Sequence 27, Appl
26	34	52.3	348	US-08-484-397A-38	Sequence 38, Appl
27	34	52.3	359	US-09-134-001C-3728	Sequence 3728, Ap

28	34	52.3	361	US-08-874-569B-21	Sequence 21, Appl
29	34	52.3	375	US-08-205-719-2	Sequence 2, Appl
30	34	52.3	375	US-08-746-883-5	Sequence 5, Appl
31	34	52.3	643	US-08-245-511-47	Sequence 47, Appl
32	34	52.3	643	US-08-600-993A-47	Sequence 47, Appl
33	34	52.3	114	US-08-637-759B-441	Sequence 441, App
34	33	50.8	114	US-08-871-355A-441	Sequence 441, App
35	33	50.8	114	US-09-201-945-441	Sequence 441, App
36	33	50.8	326	US-09-154-874-9	Sequence 9, Appl
37	33	50.8	562	US-08-427-097-30	Sequence 30, Appl
38	33	50.8	562	US-08-878-957-30	Sequence 30, Appl
39	33	50.8	564	US-08-427-097-2	Sequence 2, Appl
40	33	50.8	564	US-08-427-097-14	Sequence 14, Appl
41	33	50.8	564	US-08-427-097-16	Sequence 16, Appl
42	33	50.8	564	US-08-427-097-20	Sequence 20, Appl
43	33	50.8	564	US-08-427-097-28	Sequence 28, Appl
44	33	50.8	564	US-08-878-957-2	Sequence 2, Appl
45	33	50.8	564	US-08-878-957-14	Sequence 14, Appl

#### ALIGNMENTS

RESULT 1  
US-08-139-609-5  
Sequence 5, Application US/08139609  
Patent No. 5837249  
GENERAL INFORMATION:  
APPLICANT: Heber-Katz, Ellen  
TITLE OF INVENTION: Method for Generating an Immunogenic T  
TITLE OF INVENTION: Cell Response Protective Against a Virus  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr., P.O. Box 457  
City: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/139,609  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/868,946  
FILING DATE: 15-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/685,459  
FILING DATE: 12-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/047,443  
FILING DATE: 08-MAY-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/725,087  
FILING DATE: 19-APR-1985  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: WSTIDUSA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown

MOLECULE TYPE: peptide  
US-08-139-609-5

Query Match 63.1%; Score 41; DB 2; Length 25;  
Best Local Similarity 72.7%; Pred. No. 0.53;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VMGVLACAP 11  
: | | | | | | | | | |  
Db 13 LNAVTLACAP 23

## RESULT 2

US-08-245-511-48  
Sequence 48, Application US/08245511  
Patent No. 5928900  
GENERAL INFORMATION:  
APPLICANT: Masure, H Robert  
APPLICANT: Pearce, Barbara J  
APPLICANT: Tuomanen, Elaine  
TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND  
TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauder & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/245,511  
FILING DATE: 18-MAY-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/116,541  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-069 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 642 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
IMMEDIATE SOURCE:  
CLONE: amia  
FEATURE:  
OTHER INFORMATION: NOTE: the reference contains a  
OTHER INFORMATION: sequence error; the correct sequence shown below is obtained  
PUBLICATION INFORMATION:  
AUTHORS: Allouing, et al.  
JOURNAL: Mol. Microbiol.  
VOLUME: 4  
PAGES: 633-644  
DATE: 1990  
US-08-245-511-48

Query Match 61.5%; Score 40; DB 2; Length 642;  
Best Local Similarity 80.0%; Pred. No. 27;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 GVLACAPSK 13  
: | | | | | | | | | |  
Db 1 GVLACSSSK 10

## RESULT 3

US-08-600-993A-48  
Sequence 48, Application US/08600993A  
Patent No. 5981229  
GENERAL INFORMATION:  
APPLICANT: Masure, H Robert  
APPLICANT: Pearce, Barbara J  
APPLICANT: Tuomanen, Elaine  
TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND  
TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauder & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/600,993A  
FILING DATE: 1-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/245,511  
FILING DATE: 18-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/116,541  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-069 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 642 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
IMMEDIATE SOURCE:  
CLONE: amia  
FEATURE:  
OTHER INFORMATION: the reference contains a sequence error; the  
OTHER INFORMATION: correct sequence shown below is obtained from GENBANK  
PUBLICATION INFORMATION:  
AUTHORS: Allouing, et al.  
JOURNAL: Mol. Microbiol.  
VOLUME: 4  
PAGES: 633-644  
DATE: 1990



US-08-600-993A-48

Query Match 61.5%; Score 40; DB 2; Length 642;

Best Local Similarity 80.0%; Pred. No. 27;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 GVLAACPSK 13  
1111111111  
Db 1 GVLAACSSK 10

RESULT 4

US-08-139-609-3

Sequence 3, Application US/08139609

Patent No. 5837249

GENERAL INFORMATION:

APPLICANT: Heber-Katz, Ellen

APPLICANT: Dietzschold, Bernhard

TITLE OF INVENTION: Method for Generating an Immunogenic T

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Cntr., P.O. Box 457

CITY: Spring House

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/139,609

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/868,946

FILING DATE: 15-APR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/685,459

FILING DATE: 12-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/047,443

FILING DATE: 08-MAY-1987

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 06/725,087

FILING DATE: 19-APR-1985

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: WSTIDUSA

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

INFORMATION FOR SEQ. ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-139-609-3

Query Match 60.0%; Score 39; DB 2; Length 25;

Best Local Similarity 63.6%; Pred. No. 1.2;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNVGLAACAP 11  
1111111111  
Db 13 LNAVTVIACAP 23

RESULT 5

US-08-139-609-12

Sequence 12, Application US/08139609

Patent No. 5837249

GENERAL INFORMATION:

APPLICANT: Heber-Katz, Ellen

APPLICANT: Dietzschold, Bernhard

TITLE OF INVENTION: Method for Generating an Immunogenic T

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Cntr., P.O. Box 457

CITY: Spring House

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/139,609

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/868,946

FILING DATE: 15-APR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/685,459

FILING DATE: 12-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/047,443

FILING DATE: 08-MAY-1987

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 06/725,087

FILING DATE: 19-APR-1985

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: WSTIDUSA

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

INFORMATION FOR SEQ. ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-139-609-12

Query Match 58.5%; Score 38; DB 2; Length 26;

Best Local Similarity 70.0%; Pred. No. 1.9;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 NVGLAACAP 11  
1111111111  
Db 15 NATVIAACAP 24

RESULT 6

US-08-469-260A-83

Sequence 83, Application US/08469260A

Patent No. 6451578

GENERAL INFORMATION:

APPLICANT: JOHN N. SIMONS

APPLICANT: TAMU J. PILOT-MATIAS

APPLICANT: GEORGE J. DAWSON

APPLICANT: GEORGE G. SCHLAUDER

APPLICANT: SURESH M. DESAI

APPLICANT: THOMAS P. LEARY

APPLICANT: ANTHONY SCOTT MUEHRHOFF  
APPLICANT: JAMES C. ERKER  
APPLICANT: SHERI L. BUIJK  
APPLICANT: ISA K. MUSHAWAR  
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS  
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE  
NUMBER OF SEQUENCES: 716  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
STREET: 100 ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,260A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/424,550  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: POREMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1422 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-469-260A-83

Query Match 55.4%; Score 36; DB 4; Length 1422;  
Best Local Similarity 87.5%; Pred. No. 3.4e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 VGVLAACA 10  
|||||  
DB 783 VGVLSACA 790

RESULT 7  
US-08-469-260A-394  
Sequence 394, Application US/08469260A  
Patent No. 6451578  
GENERAL INFORMATION:  
APPLICANT: JOHN N. SIMONS  
APPLICANT: TAMI J. PILOT-MATIAS  
APPLICANT: GEORGE J. DAWSON  
APPLICANT: GEORGE G. SCHLAUDER  
APPLICANT: SURESH M. DESAI  
APPLICANT: THOMAS P. LEARY  
APPLICANT: ANTHONY SCOTT MUEHRHOFF  
APPLICANT: JAMES C. ERKER  
APPLICANT: SHERI L. BUIJK  
APPLICANT: ISA K. MUSHAWAR  
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS  
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE  
NUMBER OF SEQUENCES: 716  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
STREET: 100 ABBOTT PARK ROAD  
CITY: ABBOTT PARK

STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,260A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/424,550  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: POREMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 394:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2864 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-469-260A-394

Query Match 55.4%; Score 36; DB 4; Length 2864;  
Best Local Similarity 87.5%; Pred. No. 7.2e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 VGVLAACA 10  
|||||  
DB 1787 VGVLSACA 1794

RESULT 8  
US-09-461-697-95  
Sequence 95, Application US/09461697  
Patent No. 6277974  
GENERAL INFORMATION:  
APPLICANT: COGENT NEUROSCIENCE, Inc.  
APPLICANT: Lo, Donald C.  
APPLICANT: Barney, Shawn  
APPLICANT: Thomas, Mary Beth  
APPLICANT: Portbury, Stuart D.  
APPLICANT: Puranam, Kasturi  
APPLICANT: Katz, Lawrence C.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
TITLE OF INVENTION: CELL DEATH  
FILE REFERENCE: 10001-005-999  
CURRENT APPLICATION NUMBER: US/09/461,697  
CURRENT FILING DATE: 1999-12-14  
NUMBER OF SEQ ID NOS: 466  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 95  
LENGTH: 52  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-461-697-95

Query Match 53.8%; Score 35; DB 4; Length 52;  
Best Local Similarity 66.7%; Pred. No. 14;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 VLAACAPSK 13  
|||||  
DB 4 VLAACSPSR 12

```
RESULT 9
US-09-516-914-1
; Sequence 1, Application US/09516914
; Patent No. 6333401
; GENERAL INFORMATION:
; APPLICANT: Breinly, Sabine
; APPLICANT: Fuchs, Georg
; TITLE OF INVENTION: Phenol-Induced Proteins of Thauera aromatica
; FILE REFERENCE: BC1006 US NA
; CURRENT APPLICATION NUMBER: US/09/516,914
; EARLIER FILING DATE: 2000-03-01
; EARLIER FILING DATE: 60/122,952
; EARLIER FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Thauera aromatica
US-09-516-914-1

Query Match
Best Local Similarity 53.8%; Score 35; DB 4; Length 485;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 VGVLAACAPS 12
Db 211 LAAMAACAPS 220

RESULT 10
US-08-686-968C-227
; Sequence 227, Application US/08686968C
; Patent No. 6221361
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; FILE REFERENCE: 39119-H/JML
; CURRENT APPLICATION NUMBER: US/08/686,968C
; CURRENT FILING DATE: 1996-07-25
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 227
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Swinepox virus
US-08-686-968C-227

Query Match
Best Local Similarity 53.8%; Score 35; DB 4; Length 569;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 NVGVLAAC 9
Db 149 NIGVTAAC 156

RESULT 11
US-09-562-737-99
; Sequence 99, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 99
```

```
; LENGTH: 749
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-99

Query Match
Best Local Similarity 53.8%; Score 35; DB 4; Length 749;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 5 VLAACAPSK 13
Db 268 ILASCSPSR 276

RESULT 12
US-09-171-461-12
; Sequence 12, Application US/09171461
; Patent No. 6335016
; GENERAL INFORMATION:
; APPLICANT: Baker, Adam
; APPLICANT: Cotten, Matthew
; APPLICANT: Chioocca, Susanna
; APPLICANT: Kurzbaue, Robert
; APPLICANT: Schaffner, Gotthold
; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
; FILE REFERENCE: 0652.1800000
; CURRENT APPLICATION NUMBER: US/09/171,461
; CURRENT FILING DATE: 1999-01-12
; EARLIER APPLICATION NUMBER: PCT/EP97/01944
; EARLIER FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 942
; TYPE: PRT
; ORGANISM: CELO Virus
; FEATURE:
; OTHER INFORMATION: Position: 18289..21117 /gene: L3 /product: L3 hexon
US-09-171-461-12

Query Match
Best Local Similarity 53.8%; Score 35; DB 4; Length 942;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 VNVGVLAACAPS 12
Db 278 VNSGTRACRPN 289

RESULT 13
US-08-185-432-17
; Sequence 17, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2556 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-06-185-432-17

Query Match  
Best local Similarity 53.8%; Score 35; DB 1; Length 2556;  
Pred. No. 9.6e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 VLACAPS 12  
11111111  
DB 829 VLACAPS 836

RESULT 14  
US-06-083-590A-20  
Sequence 20; Application US/08083590A  
Patent No. 5786158  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And  
TITLE OF INVENTION: Nucleic Acids  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/083,590A  
FILING DATE: 25-JUN-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2556 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: unknown

MOLECULE TYPE: peptide  
US-08-083-590A-20

Query Match  
Best local Similarity 53.8%; Score 35; DB 1; Length 2556;  
Pred. No. 9.6e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 VLACAPS 12  
11111111  
DB 829 VLACAPS 836

RESULT 15  
US-08-532-384-20  
Sequence 20; Application US/08532384  
Patent No. 6083904  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And  
TITLE OF INVENTION: Nucleic Acids  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/532,384  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/083,590  
FILING DATE: 25-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2556 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-532-384-20

Query Match  
Best local Similarity 53.8%; Score 35; DB 3; Length 2556;  
Pred. No. 9.6e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 VLACAPS 12  
11111111  
DB 829 VLACAPS 836

Search completed: May 20, 2003, 16:44:44  
Job time : 7 secs

GenCore version 5.1.4-P5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:49:49 : Search time 6.42857 Seconds  
(without alignments)  
200.549 Million cell updates/sec

Title: US-09-869-155-18  
Perfect score: 65  
Sequence: 1 VNVGVLAACAPSK 13

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTOS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	58.5	35	10	US-09-864-761-37459
2	38	58.5	63	9	US-10-092-154-737
3	38	58.5	63	10	US-09-764-847-737
4	36	55.4	1422	8	US-08-424-550B-83
5	36	55.4	2862	10	US-09-742-659-5
6	36	55.4	2864	8	US-08-424-550B-394
7	36	55.4	2864	10	US-09-742-659-2
8	36	55.4	2865	10	US-09-742-659-6
9	35	53.8	52	10	US-09-922-261-95
10	35	53.8	485	10	US-09-870-162A-1
11	35	53.8	749	9	US-10-211-962-99
12	35	53.8	942	9	US-09-970-711-12
13	35	53.8	2444	10	US-09-944-849-2
14	34	52.3	23	9	US-10-097-065-597
15	34	52.3	135	9	US-10-097-065-596
16	34	52.3	143	9	US-09-925-299-1335
17	34	52.3	143	10	US-09-925-299-1335
18	34	52.3	210	9	US-10-072-349-90
19	34	52.3	210	10	US-09-764-855-90

20	34	52.3	258	10	US-09-864-761-37821	Sequence 37821, A
21	34	52.3	323	10	US-09-815-242-10321	Sequence 10321, A
22	34	52.3	323	10	US-09-815-242-13734	Sequence 13734, A
23	34	52.3	330	9	US-09-738-626-3767	Sequence 3767, Ap
24	34	52.3	346	9	US-09-977-577-8	Sequence 8, Appli
25	34	52.3	347	9	US-09-977-577-1	Sequence 1, Appli
26	34	52.3	347	9	US-09-977-577-3	Sequence 3, Appli
27	34	52.3	361	10	US-09-955-518-21	Sequence 21, Appli
28	34	52.3	406	9	US-09-977-577-2	Sequence 2, Appli
29	34	52.3	414	10	US-09-741-669-453	Sequence 453, App
30	34	52.3	414	10	US-09-815-242-10030	Sequence 10030, A
31	34	52.3	414	10	US-09-815-242-14077	Sequence 14077, A
32	34	52.3	431	10	US-09-925-302-485	Sequence 485, App
33	34	52.3	484	10	US-09-815-242-10457	Sequence 10457, A
34	33.5	51.5	333	10	US-09-825-882-2	Sequence 2, Appli
35	33	50.8	33	9	US-09-782-704A-1	Sequence 1, Appli
36	33	50.8	100	10	US-09-867-550-58	Sequence 58, Appli
37	33	50.8	112	9	US-10-138-516-2	Sequence 2, Appli
38	33	50.8	112	9	US-10-146-130-4	Sequence 4, Appli
39	33	50.8	112	9	US-10-092-934-3	Sequence 3, Appli
40	33	50.8	192	10	US-09-764-853-566	Sequence 566, App
41	33	50.8	377	9	US-09-978-295A-206	Sequence 206, App
42	33	50.8	377	9	US-09-978-697-206	Sequence 206, App
43	33	50.8	377	9	US-09-978-192A-206	Sequence 206, App
44	33	50.8	377	9	US-09-999-832A-206	Sequence 206, App
45	33	50.8	377	9	US-09-978-189-206	Sequence 206, App

#### ALIGNMENTS

RESULT 1  
US-09-864-761-37459  
Sequence 37459, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecolica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30

```
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 37459
;; LENGTH: 35
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC012002.2
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.9
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.3
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.8
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 5.5
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.9
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.3
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
;; OTHER INFORMATION: SWISSPROT HIT: P57072, EVALU6 6.40e+00
;; OTHER INFORMATION: EST_HUMAN HIT: BE019472.1, EVALU6 1.70e+00
US-09-864-761-37459
```

```
Query Match          58.5%; Score 38; DB 10; Length 35;
Best Local Similarity 77.8%; Pred. No. 5.7;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 4 GYLAACAPS 12
    1 1 1 1 1 1 1
Db 2 GYFACCAPS 10
```

```
RESULT 2
US-10-092-154-737
;; Sequence 737, Application US/10092154
;; Publication No. US20030054375A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC009G1
;; CURRENT APPLICATION NUMBER: US/10/092,154
;; CURRENT FILING DATE: 2002-03-07
;; NUMBER OF SEQ ID NOS: 2003
;; Prior Application removed - See File Wrapper or Palm
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 737
;; LENGTH: 63
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (32)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;; NAME/KEY: misc_feature
;; LOCATION: (44)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-092-154-737
```

```
Query Match          58.5%; Score 38; DB 9; Length 63;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 3 GYLAACAP 11
    1 1 1 1 1 1 1
Db 45 GYLAICSP 53
```

```
RESULT 3
US-09-764-847-737
;; Sequence 737, Application US/09764847
;; Patent No. US20020132767A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC009
;; CURRENT APPLICATION NUMBER: US/09/764,847
;; CURRENT FILING DATE: 2001-01-17
;; Prior application data removed - consult PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 2003
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 737
;; LENGTH: 63
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (32)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;; NAME/KEY: SITE
;; LOCATION: (44)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-847-737
```

```
Query Match          58.5%; Score 38; DB 10; Length 63;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 3 GYLAACAP 11
    1 1 1 1 1 1 1
Db 45 GYLAICSP 53
```

```
RESULT 4
US-08-424-550B-83
;; Sequence 83, Application US/08424550B
;; Patent No. US20020119447A1
;; GENERAL INFORMATION:
;; APPLICANT: JOHN N. SIMONS
;; APPLICANT: TAMM J. PILOT-MATTIAS
;; APPLICANT: GEORGE J. DAWSON
;; APPLICANT: GEORGE G. SCHLAUDER
;; APPLICANT: SUPRESE M. DESAI
;; APPLICANT: THOMAS P. LEARY
;; APPLICANT: ANTHONY SCOTT MUEHROFF
;; APPLICANT: JAMES C. ERKER
;; APPLICANT: SHERI L. BUIER
;; APPLICANT: ISA K. MUSHAWAR
;; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
;; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
;; NUMBER OF SEQUENCES: 716
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
;; STREET: 100 ABBOTT PARK ROAD
;; CITY: ABBOTT PARK
;; STATE: IL
;; COUNTRY: USA
;; ZIP: 60064-3500
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/424,550B
;; FILING DATE:
;; CLASSIFICATION: 435435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: FOREMSKI, PRISCILLA E.
;; REGISTRATION NUMBER: 33,207
```

REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1422 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-424-3508-83

Query Match 55.4%; Score 36; DB 8; Length 1422;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 VGVLAACA 10  
||||:||||  
Db 1783 VGVLSACA 1792

RESULT 5  
US-09-742-659-5  
Sequence 5, Application US/09742659  
Patent No. US20010034019A1  
GENERAL INFORMATION:  
APPLICANT: Hong, Zhi  
APPLICANT: Butkiewicz, Nancy J.  
APPLICANT: Zhong, Weidong  
APPLICANT: Ingravallo, Paul  
APPLICANT: Wright-Minogue, Jacquelyn  
APPLICANT: Lau, Johnson Y.  
APPLICANT: Lemon, Stanley M.  
TITLE OF INVENTION: Chimeric HCV/GBV-B viruses  
FILE REFERENCE: ID01116  
CURRENT APPLICATION NUMBER: US/09/742,659  
CURRENT FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: US 60/171,469  
PRIOR FILING DATE: 1999-12-22  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 5  
LENGTH: 2862  
TYPE: PRT  
ORGANISM: GBV-B/HCV  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (945)..(1129)  
OTHER INFORMATION: chimeric region  
NAME/KEY: SITE  
LOCATION: (1579)..(1593)  
OTHER INFORMATION: chimeric region  
US-09-742-659-5

Query Match 55.4%; Score 36; DB 10; Length 2862;  
Best Local Similarity 87.5%; Pred. No. 1.2e+03;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 VGVLAACA 10  
||||:||||  
Db 1785 VGVLSACA 1792

RESULT 6  
US-08-424-550B-394  
Sequence 394, Application US/08424550B  
Patent No. US20020119447A1  
GENERAL INFORMATION:  
APPLICANT: JOHN N. SIMONS  
APPLICANT: TAMM J. PILOT-MATTIAS  
APPLICANT: GEORGE J. DAWSON  
APPLICANT: GEORGE G. SCHLAUDER

APPLICANT: SURESH M. DESAI  
APPLICANT: THOMAS P. LEARY  
APPLICANT: ANTHONY SCOTT MUEHROFF  
APPLICANT: JAMES C. ERKER  
APPLICANT: SHERI L. BOJIK  
APPLICANT: ISA K. MUSHAMMAR  
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS  
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE  
NUMBER OF SEQUENCES: 716  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
STREET: 100 ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/424,550B  
APPLICATION NUMBER: US/08/424,550B  
FILING DATE:  
CLASSIFICATION: 435435  
ATTORNEY/AGENT INFORMATION:  
NAME: POREMBSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 394:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2864 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-424-550B-394

Query Match 55.4%; Score 36; DB 8; Length 2864;  
Best Local Similarity 87.5%; Pred. No. 1.2e+03;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 VGVLAACA 10  
||||:||||  
Db 1787 VGVLSACA 1794

RESULT 7  
US-09-742-659-2  
Sequence 2, Application US/09742659  
Patent No. US20010034019A1  
GENERAL INFORMATION:  
APPLICANT: Hong, Zhi  
APPLICANT: Butkiewicz, Nancy J.  
APPLICANT: Zhong, Weidong  
APPLICANT: Ingravallo, Paul  
APPLICANT: Wright-Minogue, Jacquelyn  
APPLICANT: Lau, Johnson Y.  
APPLICANT: Lemon, Stanley M.  
TITLE OF INVENTION: Chimeric HCV/GBV-B viruses  
FILE REFERENCE: ID01116  
CURRENT APPLICATION NUMBER: US/09/742,659  
CURRENT FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: US 60/171,469  
PRIOR FILING DATE: 1999-12-22  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 2  
LENGTH: 2864  
TYPE: PRT  
ORGANISM: GB virus-B

US-09-742-659-2

Query Match 55.4%; Score 36; DB 10; Length 2864;  
Best Local Similarity 87.5%; Pred. No. 1.2e+03;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 VGVLAACA 10  
11111111  
DB 1787 VGVLSACA 1794

RESULT 8

US-09-742-659-6  
Sequence 6, Application US/09742659  
Patent No. US20010034019A1  
GENERAL INFORMATION:  
APPLICANT: Hong, Zhi  
APPLICANT: Butkiewicz, Nancy J.  
APPLICANT: Zhong, Weidong  
APPLICANT: Ingravallo, Paul  
APPLICANT: Wright-Minogue, Jacquelyn  
APPLICANT: Lau, Johnson Y.  
APPLICANT: Lemon, Stanley M.  
TITLE OF INVENTION: Chimeric HCV/GBV-B viruses  
FILE REFERENCE: ID01116  
CURRENT APPLICATION NUMBER: US/09/742,659  
CURRENT FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: US 60/171,469  
PRIOR FILING DATE: 1999-12-22  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 2865  
TYPE: PRT  
ORGANISM: GBV-B/HCV  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (2275)..(2865)  
OTHER INFORMATION: chimeric region  
US-09-742-659-6

Query Match 55.4%; Score 36; DB 10; Length 2865;  
Best Local Similarity 87.5%; Pred. No. 1.2e+03;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 VGVLAACA 10  
11111111  
DB 1787 VGVLSACA 1794

RESULT 9

US-09-922-261-95  
Sequence 95, Application US/09922261  
Patent No. US2002011471A1  
GENERAL INFORMATION:  
APPLICANT: COGENT NEUROSCIENCE, Inc.  
APPLICANT: Lo, Donald C.  
APPLICANT: Barney, Shawn  
APPLICANT: Thomas, Mary Beth  
APPLICANT: Portbury, Stuart D.  
APPLICANT: Purnam, Kasturi  
APPLICANT: Katz, Lawrence C.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
TITLE OF INVENTION: CELL DEATH  
FILE REFERENCE: 10001-005-999  
CURRENT APPLICATION NUMBER: US/09/922,261  
CURRENT FILING DATE: 2001-08-03  
PRIOR APPLICATION NUMBER: US/09/461,697  
PRIOR FILING DATE: 1999-12-14  
NUMBER OF SEQ ID NOS: 466  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 95

LENGTH: 52  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-922-261-95

Query Match 53.8%; Score 35; DB 10; Length 52;  
Best Local Similarity 66.7%; Pred. No. 29;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 5 VLAACAPSK 13  
11111111  
DB 4 VLAACAPSK 12

RESULT 10

US-09-870-162A-1  
Sequence 1, Application US/09870162A  
Patent No. US20020042118A1  
GENERAL INFORMATION:  
APPLICANT: Breinig, Sabine  
APPLICANT: Fuchs, Georg  
TITLE OF INVENTION: Phenol Induced Proteins of Thauera aromatica  
FILE REFERENCE: BC1006 US DIV  
CURRENT APPLICATION NUMBER: US/09/870,162A  
CURRENT FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: 09/516914  
PRIOR FILING DATE: 2000-03-01  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 1  
LENGTH: 485  
TYPE: PRT  
ORGANISM: Thauera aromatica  
US-09-870-162A-1

Query Match 53.8%; Score 35; DB 10; Length 485;  
Best Local Similarity 60.0%; Pred. No. 2.9e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 VGVLAACAPS 12  
11111111  
DB 211 LAAMAACAPS 220

RESULT 11

US-10-211-962-99  
Sequence 99, Application US/10211962  
Publication No. US20030082640A1  
GENERAL INFORMATION:  
APPLICANT: Hertz, Joachim  
APPLICANT: Gotthardt, Michael  
TITLE OF INVENTION: LDL Receptor Signaling Pathways  
FILE REFERENCE: UTSW0708  
CURRENT APPLICATION NUMBER: US/10/211,962  
CURRENT FILING DATE: 2002-08-01  
PRIOR APPLICATION NUMBER: US/09/562,737  
PRIOR FILING DATE: 2000-05-01  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 99  
LENGTH: 749  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE: Artificial Sequence  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-10-211-962-99

Query Match 53.8%; Score 35; DB 9; Length 749;  
Best Local Similarity 55.6%; Pred. No. 4.6e+02;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 5 VLAACAPSK 13



Db 268 ILASCSPSR 276

|||||:

RESULT 12  
US-09-970-711-12

; Sequence 12, Application US/09970711  
; Patent No. US20020081279A1  
; GENERAL INFORMATION:

; APPLICANT: Baker, Adam

; APPLICANT: Cotten, Matthew

; APPLICANT: Chiocca, Susanna

; APPLICANT: Kuizbauer, Robert

; APPLICANT: Schaffner, Gotthold

; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus

; FILE REFERENCE: 0652.1800001

; CURRENT APPLICATION NUMBER: US/09/970,711

; CURRENT FILING DATE: 2001-10-05

; PRIOR APPLICATION NUMBER: 09/171,461

; PRIOR FILING DATE: 1999-01-12

; PRIOR APPLICATION NUMBER: PCT/EP97/01944

; PRIOR FILING DATE: 1997-04-18

; NUMBER OF SEQ ID NOS: 54

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 12

; LENGTH: 942

; TYPE: PRF

; ORGANISM: CELO Virus

; FEATURE: OTHER INFORMATION: Position: 18289..21117 /gene: L3 /product: L3 hexon

US-09-970-711-12

Query Match 53.8%; Score 35; DB 10; Length 942;  
Best Local Similarity 50.0%; Pred. No. 5.9e+02;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VNVGLACAPS 12

||| | | | |

Db 278 VNSGTRACRPN 289

RESULT 13

US-09-944-849-2

; Sequence 2, Application US/09944849

; Patent No. US20020151487A1

; GENERAL INFORMATION:

; APPLICANT: Nickoloff, Brian

; APPLICANT: Mele, Lucio

; TITLE OF INVENTION: METHOD AND REAGENTS FOR EPITHELIAL BARRIER FORMATION AND TREATMENT

; FILE REFERENCE: 212583

; CURRENT APPLICATION NUMBER: US/09/944,849

; CURRENT FILING DATE: 2001-08-31

; PRIOR APPLICATION NUMBER: US 60/229,614

; PRIOR FILING DATE: 2000-08-31

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 2444

; TYPE: PRF

; ORGANISM: Artificial Sequence

; FEATURE: NAME/KEY: misc\_feature

; LOCATION: (891)..(891)

; OTHER INFORMATION: The 'Xaa' at location 891 stands for Gly, or Ala.

; NAME/KEY: misc\_feature

; LOCATION: (1763)..(1763)

; OTHER INFORMATION: The 'Xaa' at location 1763 stands for Gln, Arg, Pro, or Leu.

; NAME/KEY: misc\_feature

; LOCATION: (1787)..(1787)

; OTHER INFORMATION: The 'Xaa' at location 1787 stands for Thr, Ala, Pro, or Ser.

; OTHER INFORMATION: Constitutively Active No. US20020151487A1ch-1

US-09-944-849-2

Query Match 53.8%; Score 35; DB 10; Length 2444;  
Best Local Similarity 87.5%; Pred. No. 1.6e+03;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 VLAACAPS 12

|||||

Db 829 VLAPCAPS 836

RESULT 14

US-10-097-065-597

; Sequence 597, Application US/10097065

; Publication No. US20030055236A1

; GENERAL INFORMATION:

; APPLICANT: Moore, Paul A. et al.

; TITLE OF INVENTION: 110 Human Secreted Proteins

; FILE REFERENCE: P2021P1

; CURRENT APPLICATION NUMBER: US/10/097,065

; CURRENT FILING DATE: 2002-03-14

; PRIOR APPLICATION NUMBER: PCT/US98/27059

; PRIOR FILING DATE: 1998-12-17

; PRIOR APPLICATION NUMBER: 60/070,923

; PRIOR FILING DATE: 1997-12-18

; PRIOR APPLICATION NUMBER: 60/068,007

; PRIOR FILING DATE: 1997-12-18

; PRIOR APPLICATION NUMBER: 60/068,057

; PRIOR FILING DATE: 1997-12-18

; PRIOR APPLICATION NUMBER: 60/068,006

; PRIOR FILING DATE: 1997-12-18

; PRIOR APPLICATION NUMBER: 60/068,369

; PRIOR FILING DATE: 1997-12-19

; PRIOR APPLICATION NUMBER: 60/068,367

; PRIOR FILING DATE: 1997-12-19

; PRIOR APPLICATION NUMBER: 60/068,368

; PRIOR FILING DATE: 1997-12-19

; PRIOR APPLICATION NUMBER: 60/068,169

; PRIOR FILING DATE: 1997-12-19

; PRIOR APPLICATION NUMBER: 60/068,053

; PRIOR FILING DATE: 1997-12-18

; PRIOR APPLICATION NUMBER: 60/068,064

; PRIOR FILING DATE: 1997-12-18

; PRIOR APPLICATION NUMBER: 60/068,054

; PRIOR FILING DATE: 1997-12-18

; PRIOR APPLICATION NUMBER: 60/068,008

; PRIOR FILING DATE: 1997-12-18

; PRIOR APPLICATION NUMBER: 60/068,365

; PRIOR FILING DATE: 1997-12-19

; NUMBER OF SEQ ID NOS: 672

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 597

; LENGTH: 23

; TYPE: PRF

; ORGANISM: Homo sapiens

US-10-097-065-597

Query Match 52.3%; Score 34; DB 9; Length 23;  
Best Local Similarity 60.0%; Pred. No. 19;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 VGVLAACAPS 12

:| | | | |

Db 14 IGHAVCTPS 23

RESULT 15

US-10-097-065-596

; Sequence 596, Application US/10097065

; Publication No. US20030055236A1

; GENERAL INFORMATION:

; APPLICANT: Moore, Paul A. et al.

; TITLE OF INVENTION: 110 Human Secreted Proteins

; FILE REFERENCE: P2021P1

;; CURRENT APPLICATION NUMBER: US/10/097,065  
;; CURRENT FILING DATE: 2002-03-14  
;; PRIOR APPLICATION NUMBER: PCT/US98/27059  
;; PRIOR FILING DATE: 1998-12-17  
;; PRIOR APPLICATION NUMBER: 60/070,923  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,007  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,057  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,006  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,369  
;; PRIOR FILING DATE: 1997-12-19  
;; PRIOR APPLICATION NUMBER: 60/068,367  
;; PRIOR FILING DATE: 1997-12-19  
;; PRIOR APPLICATION NUMBER: 60/068,368  
;; PRIOR FILING DATE: 1997-12-19  
;; PRIOR APPLICATION NUMBER: 60/068,169  
;; PRIOR FILING DATE: 1997-12-19  
;; PRIOR APPLICATION NUMBER: 60/068,053  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,064  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,054  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,008  
;; PRIOR FILING DATE: 1997-12-18  
;; PRIOR APPLICATION NUMBER: 60/068,365  
;; PRIOR FILING DATE: 1997-12-19  
;; NUMBER OF SEQ ID NOS: 672  
;; SOFTWARE: Patentln Ver. 2.0  
;; SEQ ID NO: 596  
;; LENGTH: 135  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-097-065-396

Query Match 52.3%; Score 34; DB 9; Length 135;  
Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 VGVYACAPS 12  
: | | | | |  
DB 32 IGHLAVCTPS 41

Search completed: May 20, 2003, 18:00:03  
Job time : 6.42857 secs

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Comphen Ltd.

# OM protein - protein search, using sw model

Run on: May 20, 2003, 16:04:12 : Search time 44.6044 Seconds  
(without alignments)  
122.483 Million cell updates/sec

Title: US-09-869-155-19

Perfect score: 225

Sequence: 1 VAMRFLCLPTGPGVAIFG.....GVPWPQFTQSNPYTLVVK 41

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

A\_Geneseq\_101002:\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	225	100.0	41	21	AA93762
2	225	100.0	41	22	AAE14683
3	225	100.0	41	22	AAU07398
4	211	93.8	370	23	AAU75816
5	211	93.8	381	23	AAU75813
6	208	92.4	381	23	AAU75812
7	206	91.6	381	23	AAU75814
8	201	89.3	370	23	AAU75815
9	195	86.7	277	23	AAU75818
10	175	77.8	185	23	AAU75824

11	175	77.8	185	23	AAU75825	Barley L endoxylan
12	175	77.8	185	23	AAU75826	Barley L endoxylan
13	114	50.7	287	23	AAU75819	Wheat L endoxylan
14	107	47.6	170	23	AAU75827	Oat L endoxylanase
15	63	28.0	693	22	AB571866	Drosophila melanog
16	62.5	27.8	62	22	AA675773	Human colon cancer
17	62	27.6	291	22	AA691308	C glutamicum prote
18	60	26.7	318	21	AB40413	Human ORFX ORF177
19	60	26.7	318	23	ABP05683	Human ORFX protein
20	60	26.7	319	22	AAU45823	Propionibacterium
21	59.5	26.4	258	21	AA642492	Arabidopsis thalia
22	59.5	26.4	258	21	AA644139	Arabidopsis thalia
23	59.5	26.4	433	21	AA642491	Arabidopsis thalia
24	59.5	26.4	433	21	AA644138	Arabidopsis thalia
25	59.5	26.4	438	21	AA642490	Arabidopsis thalia
26	59.5	26.4	439	21	AA644137	Arabidopsis thalia
27	58	25.8	92	22	AA013472	Human polypeptide
28	57.5	25.6	131	22	AB612731	Novel human diapo
29	57.5	25.6	169	22	AB612730	Novel human diapo
30	57	25.3	263	22	AB692802	Human protein sequ
31	57	25.3	279	22	AB601202	Human gene 8 encod
32	57	25.3	279	23	AB664043	Human albumin fusi
33	56.5	25.1	2114	22	AB621256	Novel human diapo
34	56	24.9	74	22	AA684155	Human Immune/haema
35	56	24.9	119	22	AAU30896	Novel human secret
36	56	24.9	196	23	AAU75817	Wheat L endoxylan
37	55.5	24.7	933	22	AB619701	Novel human diapo
38	54.5	24.2	142	22	AB612065	Novel human diapo
39	54.5	24.2	405	21	AA644494	Arabidopsis thalia
40	54.5	24.2	425	21	AA644493	Arabidopsis thalia
41	54	24.0	73	22	AB616583	Human nervous syst
42	54	24.0	95	22	AAU39802	Propionibacterium
43	54	24.0	1202	22	AB694493	Preliminary human
44	54	24.0	1221	23	AB616684	Cadherin-like asym
45	54	24.0	1221	23	AB616692	Cadherin-like asym

## ALIGNMENTS

RESULT 1

AA93762

ID AA93762 standard; peptide: 41 AA.

AC AA93762:

DT 03-OCT-2000 (first entry)

DE Amino acid sequence of an endo-beta-1,4-xylanase inhibitor peptide.

XX

XX Endo-beta-1,4-xylanase inhibitor; wheat flour; xylanase; bakery product;

KM dough; dough preparation.

XX

OS Triflicum sp.

XX

XX WO200039289-A2.

PN

XX 06-JUL-2000.

PD

XX

XX 17-DEC-1999; 99WO-IB02071;

PE

XX

XX 23-DEC-1998; 98GB-0028599.

PR

XX 06-APR-1999; 99GB-0007805.

PR

XX 15-APR-1999; 99GB-0008645.

XX

XX (DANT-) DANISCO AS.

PA

XX

XX Sildesen O, Sorensen JF;

PI

XX

XX WPI; 2000-465744/40.

DR

XX

XX Mutant xylanase protein identified using xylanase inhibitor useful for

PT preparing non-sticky dough for bakery products -

XX Claim 24; Page 112; 112pp; English.  
PS  
XX  
CC The present sequence is derived from an endo-beta-1,4-xylanase  
CC inhibitor. The protein is obtained from wheat flour. The specification  
CC also describes a mutant xylanase protein. The xylanase is useful for  
CC preparing a foodstuff, preferably a bakery product or a substance  
CC (e.g. a dough) for making the bakery product. Wild type xylanase or  
CC mutant xylanase is useful for preparing a dough that is less sticky  
CC than a dough comprising a fungal xylanase. The xylanase inhibitor is  
CC useful for screening high degree resistance xylanases for dough  
CC preparation. The xylanase is also useful for preparing a non-sticky  
CC dough. A combination of xylanase and the inhibitor is useful for  
CC calibrating and/or determining the quantity of inhibitor in a wheat  
CC flour sample.  
XX  
SQ Sequence 41 AA:  
  
Query Match 100.0%; Score 225; DB 21; Length 41;  
Best Local Similarity 100.0%; Pred. No. 3.2e-22;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VANRFLCLPTGGPGVAIFGGGVPWPQFTQSMPTLVVVK 41  
DB 1 VANRFLCLPTGGPGVAIFGGGVPWPQFTQSMPTLVVVK 41  
|||||  
RESULT 2  
AAE14693  
ID AAE14693 standard; peptide; 41 AA.  
XX  
XX AAE14693;  
AC  
XX  
XX 21-AUG-2002 (first entry)  
DT  
XX  
DE wheat flour xylanase inhibitor B chain Lys-C digested fragment #4.  
XX  
XX Refrigerated dough; syrruping; arabinoxylan; bakery product; bread;  
KW pizza base; cake; biscuit; wheat; flour; xylanase inhibitor.  
XX  
XX Trilicum aestivum.  
OS  
XX WO200152657-A1.  
PN  
XX  
PD 26-JUL-2001.  
XX  
PE 17-JAN-2001; 2001WO-1B00168.  
XX  
XX 18-JAN-2000; 2000GB-0001136.  
PR  
XX  
XX (DANI-) DANISCO AS.  
PA  
XX  
PI Poulsen CH, Sorensen JF;  
XX  
XX WPI; 2001-457446/49.  
DR  
XX  
PT Production of refrigerated dough with reduced syrruping, useful in  
PT production of bakery products such as bread, comprises admixing cereal  
PT flour, water and protein that prevents enzymatic degradation of  
PT arabinoxylan in the cereal flour -  
XX  
XX  
PS Disclosure; Page 23; 26pp; English.  
XX  
XX The invention relates to a process for producing refrigerated dough  
CC with reduced 'syrruping' (precipitation of liquid on the dough surface  
CC because of a reduction in water holding capacity caused by the breakdown  
CC of arabinoxylan over time). The process comprises admixing cereal flour  
CC and water with a protein that reduces/prevents enzymatic degradation of  
CC arabinoxylan in the cereal flour. The preferred protein is a xylanase  
CC inhibitor. The method is useful to produce refrigerated dough in which  
CC syrruping is reduced or eliminated. Refrigerated dough is typically  
CC stored for long periods to enable fresh baked products (e.g. bread,  
CC pizza bases, cakes, biscuits) to be produced as required. It may also be

CC possible to produce dough or baked products tailored to specific  
CC requirements using the method by the use of specific proteins/protein  
CC combinations. The present sequence is wheat flour  
CC endo-beta-1,4-xylanase inhibitor B chain Lys-C digested fragment.  
XX  
SQ Sequence 41 AA:  
  
Query Match 100.0%; Score 225; DB 22; Length 41;  
Best Local Similarity 100.0%; Pred. No. 3.2e-22;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VANRFLCLPTGGPGVAIFGGGVPWPQFTQSMPTLVVVK 41  
DB 1 VANRFLCLPTGGPGVAIFGGGVPWPQFTQSMPTLVVVK 41  
|||||  
RESULT 3  
AAU07398  
ID AAU07398 standard; protein; 41 AA.  
XX  
XX AAU07398;  
AC  
XX  
XX 18-DEC-2001 (first entry)  
DT  
XX  
DE Bacillus subtilis xylanase inhibitor #7.  
XX  
XX Bacillus subtilis xylanase inhibitor #7.  
DE  
XX Xylanase: plant cell wall; baking; cereal; starch production; wood;  
KW wood pulp bleaching; hemicellulose; arabinoxylan; foodstuff; inhibitor.  
XX  
XX Bacillus subtilis.  
OS  
XX WO200166711-A1.  
PN  
XX  
PD 13-SEP-2001.  
XX  
XX  
PE 08-MAR-2001; 2001WO-1B00426.  
XX  
XX 08-MAR-2000; 2000GB-0005585.  
PR  
XX 27-JUN-2000; 2000GB-0015751.  
PR  
XX  
XX (DANI-) DANISCO AS.  
PA  
XX  
PI Sibsen O, Sorensen JF;  
XX  
XX WPI; 2001-596834/67.  
DR  
XX  
PT Novel variant xylanase polypeptide or its fragment useful for degrading  
PT or modifying plant cell wall, comprises amino acid modifications such  
PT that the polypeptide has altered sensitivity to xylanase inhibitor -  
XX  
XX  
PS Disclosure; Page 63; 70pp; English.  
XX  
XX The invention relates to a variant xylanase polypeptide (I) or its  
CC fragment having xylanase activity, comprising one or more amino acid  
CC modifications such that (I) or its fragment has an altered sensitivity to  
CC a xylanase inhibitor as compared with the parent xylanase enzyme. (I) or  
CC its coding sequence (II) is useful for degrading or modifying plant cell  
CC wall or for processing a plant material by contacting the plant cell wall  
CC or plant material with (I) or (II). (I) is useful for modifying plant  
CC materials, and in baking, processing cereals, starch production,  
CC processing wood and enhancing the bleaching of wood pulp. (I)  
CC is useful for altering the viscosity derived from the presence of  
CC hemicellulose or arabinoxylan in a solution or system comprising plant  
CC cell wall material. (I) is useful for preparing a foodstuff such as  
CC bread, pretzels, tortillas, cakes, cookies, biscuits or crackers.  
CC The present sequence represents the amino acid sequence of Bacillus  
CC subtilis xylanase inhibitor #7 as described in the method of the  
CC invention.  
XX  
SQ Sequence 41 AA:  
  
Query Match 100.0%; Score 225; DB 22; Length 41;  
Best Local Similarity 100.0%; Pred. No. 3.2e-22;

Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VANNFLCLPTGGGVAIFGGGVPWPQFQSMPTLVVVK 41  
 |||  
 Db 1 VANNFLCLPTGGGVAIFGGGVPWPQFQSMPTLVVVK 41

RESULT 4  
 AAU75816  
 ID AAU75816 standard; Protein; 370 AA.  
 XX AAU75816;  
 AC  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Wheat L endoxylanase inhibitor, TAXI I, partial sequence TAXI-I.02.  
 XX  
 KW Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruiping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Triticum aestivum cultivar Estica.  
 XX  
 PN MO200198474-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 21-JUN-2001; 2001WO-BE00106.  
 XX  
 PR 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 PA  
 XX (LEUV-) LEUVEN RES & DEV.  
 PI Delcour J, Debyser W, Gebruers K, Goesaert H, Flerens K, Robben J;  
 PI Van Campenhout S;  
 DR N-PSDB; ABK13673.  
 DR  
 XX MPI; 2002-114579/15.  
 PT  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 57; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruiping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable

CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and 'noodles', animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial wheat TAXI-I ('T. aestivum L endoxylanase  
 CC inhibitor').  
 CC  
 SQ Sequence 370 AA;

Query Match 93.8%; Score 211; DB 23; Length 370;  
 Best Local Similarity 92.7%; Pred. No. 2,2e-19;  
 Matches 38; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VANNFLCLPTGGGVAIFGGGVPWPQFQSMPTLVVVK 41  
 |||  
 Db 149 VANNFLCLPTGGGVAIFGGGVPWPQFQSMPTLVVVK 189

RESULT 5  
 AAU75813  
 ID AAU75813 standard; Protein; 381 AA.  
 XX  
 AC AAU75813;  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Wheat L endoxylanase inhibitor, TAXI I, variant #2.  
 XX  
 KW Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruiping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Triticum aestivum.  
 XX  
 PN  
 XX  
 FH Key Location/Qualifiers  
 FT MISC-difference 95  
 FT MISC-difference 98 /label= Unknown  
 FT MISC-difference 101 /label= Unknown  
 FT MISC-difference 110 /label= Unknown  
 FT MISC-difference 110 /label= Unknown  
 FT MISC-difference 333 /label= Unknown  
 FT MISC-difference 333 /label= Unknown  
 PN MO200198474-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 21-JUN-2001; 2001WO-BE00106.  
 XX  
 PR 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 PA  
 XX (LEUV-) LEUVEN RES & DEV.  
 PI Delcour J, Debyser W, Gebruers K, Goesaert H, Flerens K, Robben J;  
 PI Van Campenhout S;  
 DR MPI; 2002-114579/15.

PT	Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or
PT	beta-glucanolytic enzymes comprises using endoxylanases during
PT	screening for inhibition activity or affinity chromatography with
PS	immobilised enzymes
XX	
XX	Claim 127, Page 51, 127pp; English.
CC	
CC	The invention relates to separating and/or isolating inhibitors of
CC	cellulolytic, xylanolytic and/or beta-glucanolytic enzymes compris
CC	screening the inhibition activity by using two or more enzymes during the
CC	separation and/or isolation steps that allow to distinguish inhibitors of
CC	different specificity or by using an affinity chromatographic step with
CC	immobilised enzymes and/or antibodies against inhibitors. Also
CC	included are an isolated nucleic acid molecule encoding an inhibitor
CC	which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,
CC	alpha-L-arabino-furanosidase and/or other cellulose, xylan,
CC	arabinoxyylan or beta-glucan degrading enzymes, a vector comprising the
CC	nucleic acid, an expression system transformed with the nucleic
CC	acid, a host organism transformed with the nucleic acid, the inhibitory
CC	proteins encoded by the nucleic acids and modulators of the proteins.
CC	A recombinant protein, glycoprotein or polypeptide or microorganisms,
CC	plant or plant materials transformed with the nucleic acid are useful
CC	for the formation of an endoxylanase-inhibitor complex, screening
CC	endoxylanases that are totally, less or not inhibited by the inhibitors,
CC	reducing syringing in refrigerated dough compositions, affecting the
CC	relative affinity and/or relative hydrolysis specificity and/or relative
CC	hydrolysis rate versus water-extractable and/or water-unextractable
CC	arbinoxylyans of endoxylanases such as by the formation of an
CC	endoxylanase/inhibitor complex, improving the malting of cereals such as
CC	barley, sorghum and wheat and/or the production of beer, improving the
CC	production and/or quality of baked or extruded cereal products such as
CC	straight dough, sponge dough, Chorleywood bread, breakfast cereals,
CC	biscuits, pasta and noodles, animal feed stuff, improving the production
CC	of starch derived syrups, sorbitol, xyllose and/or xylitol, wheat gluten
CC	starch separation and production, improving maize processing,
CC	plant disease resistance and nutraceutical and/or pharmaceutical
CC	applications, improving paper and pulp technologists. The present
CC	sequence is a variant of wheat TAXI-1 ('T. aestivum L endoxylanase
CC	inhibitor).
CC	Note: Variant amino acids are highlighted in the specification but
CC	no wild-type sequence is shown for comparison.
XX	
XX	Sequence 381 AA:
SO	
	Query Match 93.8%, Score 211; DB 23; Length 381;
	Best Local Similarity 92.7%; Pred. No. 2.3e-19;
	Matches 38; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY	1 VANNRLCLPTGCGPVAIFGCGPVPWQPTQSMPTLVVVK 41
DB	160 VANNRLCLPTGCGPVAIFGCGPVPWQPTQSMPTLVVTK 200
AAU75812	
AAU75812	standard; Protein: 381 AA.
AAU75812:	
23-APR-2002	(first entry)
Wheat L endoxylanase inhibitor, TAXI I, variant #1.	
Wheat: TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;	
xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;	
immobilised enzyme; enzyme; dough syringing; cereal product; pasta;	
straight dough; sponge dough; Chorleywood bread; biscuits; pasta;	
noodle; animal feed; starch separation; maize processing; malting;	
plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.	
Triticum aestivum.	

Key	Location/Qualifiers
Misc-difference 95	/label= Unknown
Misc-difference 98	/label= Unknown
Misc-difference 101	/label= Unknown
Misc-difference 110	/label= Unknown
Misc-difference 333	/label= Unknown
WO200198474-A1.	
27-DEC-2001.	
21-JUN-2001; 2001WO-BE00106.	
22-JUN-2000; 2000GB-0015296.	
25-JAN-2001; 2001GB-0002018.	
26-JAN-2001; 2001GB-0002194.	
16-MAR-2001; 2001GB-0005664.	
21-MAY-2001; 2001GB-0012328.	
(LEUV-) LEUVEN RES & DEV.	
Delcourt J, Debysier W, Gebruers K, Goesaert H, Flerens K, Robben J; Van Campenhout S;	
WPI: 2002-114579/15.	
Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or beta-glucanolytic enzymes comprises using endoxylanases during screening for inhibition activity or affinity chromatography with immobilised enzymes -	
Claim 127; Page 50-51; 127pp: English.	
The invention relates to separating and/or isolating inhibitors of cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising screening the inhibition activity by using two or more enzymes during the separation and/or isolation steps that allow to distinguish inhibitors of different specificity or by using an affinity chromatographic step with immobilised enzymes and/or antibodies against inhibitors. Also included are an isolated nucleic acid molecule encoding an inhibitor which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase, alpha-L-arabino-furanosidase and/or other cellulose, xylan, arabinoxytan or beta-glucan degrading enzymes, a vector comprising the nucleic acid, an expression system transformed with the nucleic acid, a host organism transformed with the nucleic acid, the inhibitory proteins encoded by the nucleic acids and modulators of the proteins. A recombinant protein, glycoprotein or polypeptide or microorganisms, plant or plant materials transformed with the nucleic acid are useful for the formation of an endoxylanase-inhibitor complex, screening endoxylanases that are totally, less or not inhibited by the inhibitors, reducing syruiping in refrigerated dough compositions, affecting the relative affinity and/or relative hydrolysis specificity and/or relative hydrolysis rate versus water-extractable and/or water-unextractable arabinoxytans of endoxylanases such as by the formation of an endoxylanase/inhibitor complex, improving the malting of cereals such as barley, sorghum and wheat and/or the production of beer, improving the production and/or quality of baked or extruded cereal products such as straight dough, sponge dough, Chorleywood bread, breakfast cereals, biscuits, pasta and noodles, animal feed stuff, improving the production of starch derived syrups, sorbitol, xylitol, xylitol, wheat gluten starch separation and production, improving maize processing, plant disease resistance and nutraceutical and/or pharmaceutical applications, improving paper and pulp technologies. The present sequence is a variant of wheat TAXI-I (T. aestivum L endoxylanase inhibitor).	
Note: Variant amino acids are highlighted in the specification but no wild-type sequence is shown for comparison.	

Sequence	381 AA;	92.4%;	Score 208;	DB 23;	Length 381;
Query Match		90.2%;			
Best Local Similarity		90.2%;			
Matches 37;	Conservative 2;	Mismatches 2;	Indels 0;	Gaps 0;	
OY	1 VANRFLCLPTGPGVAIFGCGPVPWQFQTSMPYTLVVK 41				
DB	160 VANRFLCLPTGPGVAIFGCGPVPWQFQTSMPYTLVVK 200				
RESULT 7					
AAU75814					
AAU75814 standard;	Protein: 381 AA.				
AAU75814;					
23-APR-2002	(first entry)				
Wheat L endoxylanase inhibitor, TAXI I, microheterogenic variant.					
Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;					
xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;					
immobilised enzyme; enzyme; dough syruuping; cereal product; beer; plant;					
straight dough; sponge dough; Chorleywood bread; biscuits; pasta;					
noodle; animal feed; starch separation; maize processing; malting;					
plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.					
Triticum aestivum.					
Key	Location/Qualifiers				
Misc-difference 95	/label= Unknown				
Misc-difference 98	/label= Unknown				
Misc-difference 101	/label= Unknown				
Misc-difference 110	/label= Unknown				
Misc-difference 145	/label= Unknown				
Misc-difference 183	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys				
Misc-difference 232	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys				
Misc-difference 275	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys				
Misc-difference 282	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys				
Misc-difference 333	/label= Asp, Asn, Val, Leu, Ser, Phe, Pro, Ala, Cys				
WO200198474-A1.					
27-DEC-2001.					
21-JUN-2001;	2001WO-BE00106.				
22-JUN-2000;	2000GB-0015296.				
25-JAN-2001;	2001GB-0002018.				
26-JAN-2001;	2001GB-0002194.				
16-MAR-2001;	2001GB-0006564.				
21-MAY-2001;	2001GB-0012328.				
(LEUV-) LEUVEN RES & DEV.					
Delcours J, Debysse W, Gebruers K, Goessaert H, Flierens K, Robben J, Van Campenhout S,					
WPI; 2001-114579/15.					
Separating and/or isolating inhibitors of cellulolytic, xyloxytic, or beta-glucanolytic enzymes comprises using endoxylanases during					

PT screening for inhibition activity or affinity chromatography with  
XX Immobilised enzymes -  
PS Claim 127; Page 51; 127pp; English.  
XX  
CC The invention relates to separating and/or isolating inhibitors of  
CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
CC screening the inhibition activity by using two, or more enzymes during the  
CC separation and/or isolation steps that allow to distinguish inhibitors of  
CC different specificity or by using an affinity chromatographic step with  
CC immobilised enzymes and/or antibodies against inhibitors. Also  
CC included are an isolated nucleic acid molecule encoding an inhibitor  
CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
CC arabinoxylian or beta-glucan degrading enzymes, a vector comprising the  
CC nucleic acid, an expression system transformed with the nucleic  
CC acid, a host organism transformed with the nucleic acid, the inhibitory  
CC proteins encoded by the nucleic acids and modulators of the proteins.  
CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
CC plant or plant materials transformed with the nucleic acid are useful  
CC for the formation of an endoxylanase-inhibitor complex, screening  
CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
CC reducing syruping in refrigerated dough compositions, affecting the  
CC relative affinity and/or relative hydrolysis specificity and/or relative  
CC hydrolysis rate versus water-extractable and/or water-unextractable  
CC arabinoxylians of endoxylanases such as by the formation of an  
CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
CC barley, sorghum and wheat and/or the production of beer, improving the  
CC production and/or quality of baked or extruded cereal products such as  
CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
CC biscuits, pasta and noodles, animal feed stuff, improving the production  
CC of starch derived syrups, sorbitol, xylitol, maltitol, wheat gluten  
CC starch separation and production, improving maize processing,  
CC plant disease resistance and nutraceutical and/or pharmaceutical  
CC applications, a microtiterogenic variant of wheat TAXI-I (T. aestivum L  
CC endoxylanase inhibitor).  
CC Note: Variant amino acids are highlighted in the specification but  
CC no wild-type sequence is shown for comparison.  
XX  
SQ Sequence 381 AA:  
  
Query Match 91.6%; Score 206; DB 23; Length 381;  
Best Local Similarity 90.2%; Pred. No. 1e-18;  
Matches 37; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
1 VANRFLLCPTGGCGVATFGGGPVWMPQFTGSMPTLVYVK 41  
II  
Db 160 VANRFLLCPTGGCGVATFGGGPVWMPQFTGSMPTLVYVK 200  
  
RESULT 8  
AAU75815  
ID: AAU75815 standard; Protein: 370 AA.  
XX  
XX AAU75815;  
XX  
XX 23-APR-2002 (first entry)  
XX  
DE Wheat L endoxylanase inhibitor, TAXI I, partial sequence TAXI-I.0.  
XX  
XX Wheat; TAXI-I; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
KW noodle; animal feed; starch separation; maize processing; malting;  
XX plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
XX  
XX Triticum aestivum cultivar Solissons.  
XX  
XX WO200198474-A1.  
XX  
XX 27-DEC-2001.

21-JUN-2001; 2001MO-BE00106.  
 22-JUN-2000; 2000GB-0015296.  
 25-JAN-2001; 2001GB-0002018.  
 26-JAN-2001; 2001GB-0002194.  
 16-MAR-2001; 2001GB-0006564.  
 21-MAY-2001; 2001GB-0012328.  
 (LEUV-) LEUVEN RES & DEV.  
 Delcour J, Debysier W, Gebruers K, Goesaert H, Fierens K, Robben J,  
 van Campenhout S;  
 WPI: 2002-114579/15.  
 R-PSDB: ABK13672.  
 Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 beta-glucanolytic enzymes comprises using endoxylanases during  
 screening for inhibition activity or affinity chromatography with  
 immobilised enzymes -  
 Claim 127; page 57; 127pp; English.  
 The invention relates to separating and/or isolating inhibitors of  
 cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 screening the inhibition activity by using two or more enzymes during the  
 separation and/or isolation steps that allow to distinguish inhibitors of  
 different specificity or by using an affinity chromatographic step with  
 immobilised enzymes and/or antibodies against inhibitors. Also  
 included are an isolated nucleic acid molecule encoding an inhibitor  
 which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 nucleic acid, an expression system transformed with the nucleic  
 acid, a host organism transformed with the nucleic acid, the inhibitory  
 proteins encoded by the nucleic acids and modulators of the proteins.  
 A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 plant or plant materials transformed with the nucleic acid are useful  
 for the formation of an endoxylanase-inhibitor complex, screening  
 endoxylanases that are totally, less or not inhibited by the inhibitors,  
 reducing syruiping in refrigerated dough compositions, affecting the  
 relative affinity and/or relative hydrolysis specificity and/or relative  
 hydrolysis rate versus water-extractable and/or water-unextractable  
 arabinoxylans of endoxylanases such as by the formation of an  
 endoxylanase/inhibitor complex. Improving the malting of cereals such as  
 barley, sorghum and wheat and/or the production of beer, improving the  
 production and/or quality of baked or extruded cereal products such as  
 straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 biscuits, pasta and noodles, animal feed stuff, improving the production  
 of starch derived syrups, sorbitol, xylitol, wheat gluten  
 starch separation and production, improving maize processing,  
 plant disease resistance and nutraceutical and/or pharmaceutical  
 applications, improving paper and pulp technologies. The present  
 sequence represents a partial wheat TAxI-1 (T. aestivum L endoxylanase  
 inhibitor).  
 Sequence 370 AA;  
 Query Match 89.3%; Score 201; DB 23; Length 370;  
 Best Local Similarity 90.2%; Pred. No. 4.5e-18;  
 Matches 37; Conservative 1; Mismatches 3; Indels 0; Gaps 0.  
 1 VANRFLCLPTGGGVAIFGGGPPVMPQFTQSMPYTLVVK 41  
 |||  
 149 VANRFLCLPTGGGLGVAIFGGGPPVMPQFTQSMPYTLVVK 189  
 RESULT 9  
 AAU75818  
 AAU75818 standard: Protein: 277 AA.  
 AAU75818:

XX	23-APR-2002	(first entry)
DT		
DE		L endoxylanase inhibitor, AXYI-II, partial sequence AXYI-II.01.
XX		
XX		AXYI-II; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;
KW		xylanolytic enzyme; inhibitor; beta-glucanolytic enzyme; inhibitor; ds;
KW		immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;
KW		straight dough; sponge dough; Chorleywood bread; biscuits; pasta;
KW		noodle; animal feed; starch separation; maize processing; malting;
XX		plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.
XX		
OS		Aegilops tauschii variety Acc2220051.
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference	4
FT		/note= "Encoded by CTC"
XX		
PN		WO200198474-A1.
XX		
PD		27-DEC-2001.
XX		
XX		21-JUN-2001; 2001WO-BE00106.
PF		
XX		
PR		22-JUN-2000; 2000GB-0015295.
PR		25-JAN-2001; 2001GB-0002018.
PR		26-JAN-2001; 2001GB-0002194.
PR		16-MAR-2001; 2001GB-0006564.
PR		21-MAY-2001; 2001GB-0012328.
XX		
PA		(LEUV-) LEUVEN RES & DEV.
XX		
P1	Delcour J, Debysse W, Gebruers K, Goesaert H, Flerens K, Robben J;	
P1	Van Campenhout S;	
PT		WPI: 2002-114579/15.
DR		N-PSDB: ABRK3675.
XX		
PT		Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or
PT		beta-glucanolytic enzymes comprises using endoxylanases during
PT		screening for inhibition activity or affinity chromatography with
PT		immobilised enzymes
XX		
PS		Claim 127; Page 58; 127pp; English.
XX		
CC		The invention relates to separating and/or isolating inhibitors of
CC		cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises
CC		screening the inhibition activity by using two or more enzymes during the
CC		separation and/or isolation steps that allow to distinguish inhibitors of
CC		different specificity or by using an affinity chromatographic step with
CC		immobilised enzymes and/or antibodies against inhibitors. Also
CC		included are an isolated nucleic acid molecule encoding an inhibitor
CC		which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,
CC		alpha-L-arabino-furanosidase and/or other cellulose, xylan,
CC		arabinoxylan or beta-glucan degrading enzymes, a vector comprising the
CC		nucleic acid, an expression system transformed with the nucleic
CC		acid, a host organism transformed with the nucleic acid, the inhibitory
CC		proteins encoded by the nucleic acids and modulators of the proteins.
CC		A recombinant protein, glycoprotein or polypeptide or microorganisms,
CC		plant or plant materials transformed with the nucleic acid are useful
CC		for the formation of an endoxylanase-inhibitor complex, screening
CC		endoxylanases that are totally, less or not inhibited by the inhibitors,
CC		reducing syruping in refrigerated dough compositions, affecting the
CC		relative affinity and/or relative hydrolysis specificity and/or relative
CC		hydrolysis rate versus water-extractable and/or water-unextractable
CC		arabinoxylans of endoxylanases such as by the formation of an
CC		endoxylanase/inhibitor complex, improving the melting of cereals such as
CC		barley, sorghum and wheat and/or the production of beer, improving the
CC		production and/or quality of baked or extruded cereal products such as
CC		straight dough, sponge dough, Chorleywood bread, breakfast cereals,
CC		biscuits, pasta and noodles, animal feed stuff, improving the production
CC		of starch derived syrups, sobitol, xylitol and/or xylool, wheat gluten
CC		starch separation and production, improving maize processing,



CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial Aegilops tauschii HVXI-II (Aegilops  
 CC tauschii L endoxylanase inhibitor).

XX  
 SQ Sequence 277 AA:

Query Match 86.7%; Score 195; DB 23; Length 277;  
 Best Local Similarity 82.9%; Pred. No. 2e-17;  
 Matches 34; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 VANRFLCLPTGGVAFIIFGGVPPWPFQTSMPYTLVVK 41  
 ||:|||||:|||||:|||||:|||||:|  
 Db 149 VANKFLCLPISGPGVAFIFGGGPLPWPQLTQSMPTPLVTK 189

RESULT 10

AAU75824  
 ID AAU75824 standard; Protein: 185 AA.

AC AAU75824;

XX 23-APR-2002 (first entry)

DE Barley L endoxylanase inhibitor, HVXI I, variant #1.

XX Barley: HVXI-I: L endoxylanase inhibitor; cellulolytic enzyme inhibitor;

KM xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;

KM immobilised enzyme; enzyme; dough syripping; cereal product; beer; plant;

KM straight dough; sponge dough; Chorleywood bread; biscuits; pasta;

KM noodle; animal feed; starch separation; maize processing; malting;

KM plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syripping in rehydrated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a variant of barley HVXI-I (H. vulgare L endoxylanase  
 CC inhibitor).

CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.

SQ Sequence 185 AA:

Query Match 77.8%; Score 175; DB 23; Length 185;  
 Best Local Similarity 78.0%; Pred. No. 5.4e-15;  
 Matches 32; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 1 VANRFLCLPTGGVAFIIFGGVPPWPFQTSMPYTLVVK 41  
 ||:|||||:|||||:|||||:|||||:|  
 Db 20 VSHRFLCLPTGGVAFIIFGGGPLPWPQLTQSMPTPLVVK 60

RESULT 11

AAU75825  
 ID AAU75825 standard; Protein: 185 AA.

AC AAU75825;

XX 23-APR-2002 (first entry)

DE Barley L endoxylanase inhibitor, HVXI I, variant #2.

XX Barley: HVXI-I: L endoxylanase inhibitor; cellulolytic enzyme inhibitor;

KM xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;

KM immobilised enzyme; enzyme; dough syripping; cereal product; beer; plant;

KM straight dough; sponge dough; Chorleywood bread; biscuits; pasta;

KM noodle; animal feed; starch separation; maize processing; malting;

KM plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

XX Hordeum vulgare.

The invention relates to separating and/or isolating inhibitors of  
 cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 screening the inhibition activity by using two or more enzymes during the  
 separation and/or isolation steps that allow to distinguish inhibitors of  
 different specificity or by using an affinity chromatographic step with  
 immobilised enzymes and/or antibodies against inhibitors. Also

PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 XX (LEUV-) LEUVEN RES & DEV.  
 PI Delcour J, Debyser W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR WFI; 2002-114579/15.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 63; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutritional and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence is a variant of barley HvXI-1 (H. vulgare L endoxylanase  
 CC inhibitor).  
 CC Note: Variant amino acids are highlighted in the specification but  
 CC no wild-type sequence is shown for comparison.  
 XX  
 SO Sequence 185 AA;  
 Query Match 77.8%; Score 175; DB 23; Length 185;  
 Best Local Similarity 78.0%; Pred. No. 5.4e-15;  
 Matches 32; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 VANKRLCLPTGGPGVAIFGGGVPMPOFTQSPMYTLVVK 41  
 DB 20 VSHRFLCLPTGGAGVAILGGGLPMPPOFTQSMAYTPLYVK 60  
 RESULT 12  
 AAU75326  
 ID AAU75826 standard; Protein; 185 AA.  
 XX  
 AC AAU75826;  
 XX  
 XX 23-APR-2002 (first entry)  
 DT  
 XX  
 DE Barley L endoxylanase inhibitor, HvXI 1, microheterogenic variant.

XX  
 XX Barley; HvXI-1; L endoxylanase inhibitor; cellulolytic enzyme inhibitor;  
 KW xylanolytic enzyme inhibitor; beta-glucanolytic enzyme inhibitor;  
 KW immobilised enzyme; enzyme; dough syruping; cereal product; beer; plant;  
 KW straight dough; sponge dough; Chorleywood bread; biscuits; pasta;  
 KW noodle; animal feed; starch separation; maize processing; malting;  
 KW plant disease resistance; nutraceutical; pharmaceutical; paper; pulp.  
 XX  
 OS Hordeum vulgare.  
 OS Synthetic.  
 XX  
 XX  
 EH Key Location/Qualifiers  
 FT MISC-difference 72  
 FT /label= Unknown  
 FT MISC-difference 128  
 FT /label= Pro, Ala, Asp, Leu, Gly, Asn, Cys  
 FT MISC-difference 132  
 FT /label= Pro, Ala, Asp, Leu, Gly, Asn, Cys  
 FT MISC-difference 134  
 FT /label= Pro, Ala, Asp, Leu, Gly, Asn, Cys  
 FT MISC-difference 148  
 FT /label= Pro, Ala, Asp, Leu, Gly, Asn, Cys  
 FT MISC-difference 165  
 FT /label= Unknown  
 PN  
 PD WO200198474-A1.  
 XX  
 XX 27-DEC-2001.  
 XX  
 XX 21-JUN-2001; 2001WO-BE00106.  
 XX  
 XX 22-JUN-2000; 2000GB-0015296.  
 PR 25-JAN-2001; 2001GB-0002018.  
 PR 26-JAN-2001; 2001GB-0002194.  
 PR 16-MAR-2001; 2001GB-0006564.  
 PR 21-MAY-2001; 2001GB-0012328.  
 XX  
 XX (LEUV-) LEUVEN RES & DEV.  
 PA  
 PI Delcour J, Debyser W, Gebruers K, Goesaert H, Fierens K, Robben J;  
 PI Van Campenhout S;  
 DR WFI; 2002-114579/15.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 63; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprises  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as



XX Delcourt J, Debysse W, Gebruers K, Goesaert H, Flereus K, Robben J;  
 PI Van Campenhout S;  
 XX  
 DR WPI: 2002-114579/15.  
 DR N-PSDB: ABK13666.  
 XX  
 PT Separating and/or isolating inhibitors of cellulolytic, xylanolytic, or  
 PT beta-glucanolytic enzymes comprises using endoxylanases during  
 PT screening for inhibition activity or affinity chromatography with  
 PT immobilised enzymes  
 XX  
 PS Claim 127; Page 65; 127pp; English.  
 XX  
 CC The invention relates to separating and/or isolating inhibitors of  
 CC cellulolytic, xylanolytic and/or beta-glucanolytic enzymes comprising  
 CC screening the inhibition activity by using two or more enzymes during the  
 CC separation and/or isolation steps that allow to distinguish inhibitors of  
 CC different specificity or by using an affinity chromatographic step with  
 CC immobilised enzymes and/or antibodies against inhibitors. Also  
 CC included are an isolated nucleic acid molecule encoding an inhibitor  
 CC which inhibits cellulase, endoxylanase, beta-glucanase, beta-xylosidase,  
 CC alpha-L-arabino-furanosidase and/or other cellulose, xylan,  
 CC arabinoxylan or beta-glucan degrading enzymes, a vector comprising the  
 CC nucleic acid, an expression system transformed with the nucleic  
 CC acid, a host organism transformed with the nucleic acid, the inhibitory  
 CC proteins encoded by the nucleic acids and modulators of the proteins.  
 CC A recombinant protein, glycoprotein or polypeptide or microorganisms,  
 CC plant or plant materials transformed with the nucleic acid are useful  
 CC for the formation of an endoxylanase-inhibitor complex, screening  
 CC endoxylanases that are totally, less or not inhibited by the inhibitors,  
 CC reducing syneruping in refrigerated dough compositions, affecting the  
 CC relative affinity and/or relative hydrolysis specificity and/or relative  
 CC hydrolysis rate versus water-extractable and/or water-unextractable  
 CC arabinoxylans of endoxylanases such as by the formation of an  
 CC endoxylanase/inhibitor complex, improving the malting of cereals such as  
 CC barley, sorghum and wheat and/or the production of beer, improving the  
 CC production and/or quality of baked or extruded cereal products such as  
 CC straight dough, sponge dough, Chorleywood bread, breakfast cereals,  
 CC biscuits, pasta and noodles, animal feed stuff, improving the production  
 CC of starch derived syrups, sorbitol, xylose and/or xylitol, wheat gluten  
 CC starch separation and production, improving maize processing,  
 CC plant disease resistance and nutraceutical and/or pharmaceutical  
 CC applications, improving paper and pulp technologies. The present  
 CC sequence represents a partial oat AsxI-I (A. sativa L endoxylanase  
 CC inhibitor).  
 CC  
 XX  
 SO Sequence 170 AA:  
 Query Match 47.6%; Score 107; DB 23; Length 170;  
 Best Local Similarity 51.0%; Pred. No. 3.7e-06;  
 Matches 25; Conservative 5; Mismatches 11; Indels 8; Gaps 2;  
 OY 1 VAKRFLCLPTG---GGGVAIFGGGPV-----PMPQFTQSMPTLVVYK 41  
 II :||||| : | ||||| : : ||| : :  
 DB 6 VAKRFLCLSRGGVGVGVAIFGGGPLHTAOPETDYTOSELYTPPTK 54  
 II :||||| : | ||||| : : ||| : :  
 RESULT 15  
 ID ABB71966 standard; Protein; 693 AA.  
 AC ABB71966;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 42690.  
 XX  
 KW Drosophila: developmental biology; cell signalling; insecticide;  
 XX pharmaceutical.  
 OS Drosophila melanogaster.  
 XX

PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR N-PSDB: ABL16069.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions  
 XX  
 PS Disclosure; SEQ ID NO 42690; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16175-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SO Sequence 693 AA:  
 Query Match 28.0%; Score 63; DB 22; Length 693;  
 Best Local Similarity 41.2%; Pred. No. 9.1;  
 Matches 14; Conservative 7; Mismatches 9; Indels 4; Gaps 2;  
 OY 7 LCUPTRGGVAIFGGGPVMPQFTQSMPTLVVY 40  
 : ||| : : || : : || : : || : :  
 DB 141 MALPSGGVGVGVC---PYPSNEON-PYPSIGV 170  
 : ||| : : || : : || : : || : :  
 Search completed: May 20, 2003, 16:14:44  
 Job time : 45.6044 secs



OY 6 LILCLPTGPGVAIFGGPVPWPQ 28  
 :| ||| | ||| |||  
 Db 184 VLSTPTGTAATASAGPVPWPE 206

# RESULT 3

succinylarginine dihydrolase [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus  
 C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001

C:Accession: G87448  
 R:Niezman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Taub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98: 4136-4141, 2001

A:Title: Complete Genome Sequence of *Caulobacter crescentus*.

A:Reference number: AB7249; MUID:21173698; PMID:11259647

A:Accession: G87448

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-446 <STO>

A:Cross-references: GB:AE005673; NID:q13423007; PIDN:AAK23587.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC1608

C:Superfamily: Escherichia coli hypothetical protein b1745

Query Match 27.1%; Score 61; DB 2; Length 446;  
 Best Local Similarity 39.4%; Pred. No. 4.5;  
 Matches 13; Conservative 2; Mismatches 18; Indels 0; Gaps 0;

# RESULT 4

hypothetical protein EC52489 [imported] - Escherichia coli (strain O157:H7, substrain R1

C:Species: Escherichia coli  
 C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C:Accession: A99940  
 R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: A99940

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-301 <HAV>

A:Cross-references: GB:BA000007; PIDN:BAH35912.1; PID:q13361956; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:

A:Gene: EC52489

C:Superfamily: Arabidopsis thaliana hypothetical protein F9D16.200

Query Match 26.4%; Score 59.5; DB 2; Length 301;  
 Best Local Similarity 38.9%; Pred. No. 4.7;  
 Matches 14; Conservative 4; Mismatches 5; Indels 13; Gaps 1;

# RESULT 5

hypothetical protein yead [imported] - Escherichia coli (strain O157:H7, substrain EDL93

C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: E85788  
 R:Perrin, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimlantia, E.; Potamousis, K.; Apoda  
 Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551

A:Accession: E85788

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-301 <STO>

A:Cross-references: GB:AE005174; NID:q12515814; PIDN:AA656769.1; GSPDB:GN00145; UNCP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: yead

C:Superfamily: Arabidopsis thaliana hypothetical protein F9D16.200

# RESULT 6

hypothetical protein b1780 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: D64938

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.: Rose, D.J.; Mau, B.; Sho, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617; PMID:9276503

A:Accession: D64938

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-301 <BLAT>

A:Cross-references: GB:AE000273; GB:U00096; NID:q1788078; PIDN:AAK74850.1; PID:q17880

A:Experimental source: strain K-12, substrain MG1655

C:Superfamily: Arabidopsis thaliana hypothetical protein F9D16.200

# RESULT 7

hypothetical protein F15K9.17 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001

C:Accession: F66163

R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Keul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

nansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: AB6141; MUID:21016719; PMID:11130712



ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G86163

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-434 <STO>

A:Cross-references: GB:AE005172; NID:g3850580; PIDN:AACT2120.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: conglutinin gamma

Query Match

Best Local Similarity 26.2%; Score 59; DB 2; Length 434;

Matches 15; Conservative 7; Mismatches 13; Indels 4; Gaps 2;

OY 4 RFLCLPTGGPVAIFGCGP---VPMPOFTQSMPTLVV 39

Db 200 KFAVCL-TSGRGVAFGNGPVPFLGIQISRLQKTPLLI 237

RESULT 13

F70073

hypothetical protein yxce - *Bacillus subtilis*

C:Species: *Bacillus subtilis*

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999

C:Accession: F70073

R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte

C.; Brun, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emmerson, P.T.; Entlan, K.D.; Errington, J.; Fabre, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall

lechi, C.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Huilo, M.F.

Koetter, P.; Konungstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lindrois,

A:Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadai, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schreier, R.; Scrofano, F.; Sekiguchi, J.; Sekowska, A.; Ser

akch, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Walters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033; PMID:9344377

A:Accession: F70073

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-216 <RUN>

A:Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CAB16015.1; PID:el184704;

A:Experimental source: strain 168

C:Genetics:

A:Gene: yxce

Query Match

Best Local Similarity 25.8%; Score 58; DB 2; Length 216;

Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 10 PTGGPVAIFGCGPVPV 26

Db 169 PVGPGVYSGASPIAW 185

RESULT 14

I38855

activator of RNA decay - human

C:Species: *Homo sapiens* (man)

C:Date: 01-Nov-1996 #sequence\_revision 01-Nov-1996 #text\_change 21-Jul-2000

C:Accession: I38856

R:Wang, M.; Cohen, S.N.

Proc. Natl. Acad. Sci. U.S.A. 91, 10591-10595, 1994

A:Title: ard-1, a human gene that reverses the effects of temperature-sensitive and dele

A:Reference number: I38856; MUID:95024160; PMID:7524097

A:Accession: I38856

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-127 <RES>

A:Cross-references: EMBL:U04575; NID:g559771; PIDN:AAA64749.1; PID:g559772

C:Genetics:

A:Gene: ard-1

Query Match

Best Local Similarity 25.3%; Score 57; DB 2; Length 127;

Matches 12; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

OY 10 PTGGPVAIFGCGPVPVPMPOFTQSMPTLVV 39

Db 54 PHGHTGTLTGGLPMYPVNLAPVDVLPVV 83

RESULT 15

I46033

NIPP-1, nuclear inhibitor of protein phosphatase-1 - bovine

C:Species: *Bos primigenius taurus* (cattle)

C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 05-Nov-1999

C:Accession: I46033

R:Van Eynde, A.; Wera, S.; Beullens, M.; Torrekens, S.; van Leuven, F.; Stalmans, W.;

J. Biol. Chem. 270, 28068-28074, 1995

A:Title: Molecular cloning of NIPP-1, a nuclear inhibitor of protein phosphatase-1, r

A:Reference number: I46033; MUID:96081835; PMID:7499293

A:Accession: I46033

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-351 <VAN>

A:Cross-references: EMBL:250748; NID:g1082085; PIDN:CAA90625.1; PID:g1082086

Query Match

Best Local Similarity 25.3%; Score 57; DB 2; Length 351;

Matches 12; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

OY 10 PTGGPVAIFGCGPVPVPMPOFTQSMPTLVV 39

Db 278 PHGHTGTLTGGLPMYPVNLAPVDVLPVV 307

Search completed: May 20, 2003, 16:37:08  
Job time : 18.6703 secs



GenCore version 5.1.4-P5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:05:42 ; Search time 11.0385 Seconds

(without alignments)  
154.055 Million cell updates/sec

Title: US-09-869-155-19

Perfect score: 225  
Sequence: 1 VANRFLCLPTGPGVAIFG.....GVPWPQFTQSMPTLVVVK 41

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	28.0	693	1 CAUP_DROME	P54269 drosophila
2	62.5	27.8	274	1 PPNK_AQUAE	O67055 aquifex aeo
3	62	27.6	301	1 PPNK_STRCO	Q9s219 streptomyces
4	59.5	26.4	294	1 UP08_ECOCI	P39173 escherichia
5	59	26.2	231	1 CHAC_ECOCI	P39163 escherichia
6	59	26.2	307	1 PPNK_MYCTU	O33196 mycobacteri
7	59	26.2	311	1 PPNK_MYCTU	O49897 mycobacteri
8	58	25.6	216	1 YXCE_BACSU	P46335 bacillus su
9	57.5	25.3	351	1 PPR8_BOVIN	O28147 bos taurus
10	57	25.3	351	1 PPR8_BOVIN	O12972 homo sapien
11	57	25.3	314	1 COBD_RHOCA	O52683 rhodobacter
12	53	23.6	699	1 CRAR_HUMAN	P48740 h complemen
13	53	23.6	356	1 IRXH_CAEEL	O91444 caenorhabdl
14	52.5	23.3	441	1 PAC1_MOUSE	O61644 mus musculu
15	52.5	23.3	441	1 PAC1_MOUSE	O920w5 r protein k
16	52.5	23.3	441	1 PAC1_MOUSE	O9b111 homo sapien
17	52.5	23.3	444	1 PAC1_MOUSE	P37122 solanum mel
18	52.5	23.3	505	1 C762_SOLME	O9cnu2 pasteurella
19	51.5	22.9	305	1 PPNK_PASUO	O03263 saccharomyc
20	51.5	22.9	567	1 YN8M_YEAST	O96449 sus sapien
21	51.5	22.9	567	1 YN8M_YEAST	O9bnc0 pseudomonas
22	51	22.7	452	1 TPEE_DROME	P48613 drosophila
23	51	22.7	452	1 TPEE_DROME	O9w84 mus musculu
24	50.5	22.4	186	1 NDKM_MOUSE	O9w84 mus musculu
25	50.5	22.4	205	1 RACU_DICDI	O9w84 mus musculu
26	50.5	22.4	294	1 PPNK_VIBCH	O9w84 mus musculu
27	50.5	22.4	423	1 GRPR_PIG	P34399 sus scrofa
28	50.5	22.4	504	1 CPDI_RAT	P10633 rattus norv
29	50.5	22.4	504	1 CPDI_RAT	P24456 mus musculu
30	50.5	22.4	505	1 CPDI_MOUSE	P24457 mus musculu
31	50	22.2	166	1 STEE_ECOLI	P33227 escherichia
32	50	22.2	236	1 SLR2_RALSO	O30919 ralsstonia s
33	50	22.2	329	1 Y461_CHLTR	O84467 chlamydia t

34	50	22.2	329	1 Y746_CHLMO	Q9p122 chlamydia m
35	50	22.2	354	1 FXE1_HUMAN	Q12946 homo sapien
36	50	22.2	454	1 AATM_LUPAN	P26563 lupinus ang
37	50	22.2	636	1 P73_HUMAN	O15550 homo sapien
38	50	22.2	693	1 S6A3_BOVIN	P27922 bos taurus
39	50	22.2	1586	1 SN22_HUMAN	P51531 homo sapien
40	49.5	22.0	305	1 PPN1_SYNY3	P74430 synecocyst
41	49.5	22.0	487	1 XYLC_PSEPU	P43503 pseudomonas
42	49.5	22.0	504	1 CPD9_MOUSE	P11714 mus musculu
43	49.5	22.0	897	1 CYRB_HUMAN	P32927 homo sapien
44	49	21.8	277	1 PPNK_PPRAB	O9v081 pyrococcus
45	49	21.8	292	1 PPNK_BUCAL	P57282 buchnera ap

## ALIGNMENTS

CAUP_DROME	STANDARD:	PRT:	693 AA.
ID	CAUP_DROME	STANDARD:	PRT:
AC	P54269: Q9VU00:		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Homeobox protein caupolican.		
GN	CAUP OR CG10605.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;		
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96180722; PubMed=8620542;		
RA	Gomez-Skarmeta J.-L., del Corral R.D., de la Calle-Mustienes E.,		
RA	Ferres-Marco D., Modolell J.;		
RT	*Araucan and caupolican, two members of the novel iroquois complex,		
RT	RT encode homeoproteins that control proneural and vein-forming genes.;		
RL	Cell 85:95-110(1996).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Berkely;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Manamides P.C., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abrill J.F., Abmayyan A., An H.-J., Andrews-Plantkooch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Bertman B.P., Bhattacharya S.,		
RA	Bortolova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doop L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,		
RA	Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,		
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,		
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Keichum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Merkllov G., Mlshina N.V., Moberly C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclied J.M.,		
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,		
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,		
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,		

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao O.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Glisb R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 Science 287:2185-2195(2000).  
 CC -1- FUNCTION: CONTROL PRONEURAL AND VEIN FORMING GENES. POSITIVE  
 CC TRANSCRIPTIONAL CONTROLLER OF AC-SC (ACHAETE-SCURF). MAY ACT AS AN  
 CC ACTIVATOR THAT INTERACTS WITH THE TRANSCRIPTIONAL COMPLEX  
 CC ASSEMBLED ON THE AC AND SC PROMOTERS AND PARTICIPATES IN  
 CC TRANSCRIPTION INITIATION.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- MISCELLANEOUS: 'CAPOLOCAN' IS NAMED AFTER THE ARAUCANIAN  
 CC AMERICAN-INDIAN TRIBE. ALSO CALLED MOHAWKS, WHO SHAVED ALL BUT A  
 CC MEDIAL STRIPE OF HAIRS ON THE HEAD.  
 CC -1- SIMILARITY: BELONGS TO THE TALE/IR0 HOMEBOX FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X95178; CAA64485.1; -  
 CC EMBL: AE003540; AAF49895.1; -  
 CC HSP: P41778; IDU6.  
 DR FlyBase: FBgn0015919; caup.  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR003893; Iroquois\_homeo.  
 DR Pfam: PF00046; homeobox; 1.  
 DR ProDom: PD000010; Homeobox; 1.  
 DR SMART: SM00389; HOX; 1.  
 DR SMART: SM00548; IRO; 1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 KM Transcriptional regulation; DNA-binding; Homeobox; Nuclear protein;  
 KM Developmental protein.  
 FT DNA\_BIND 226 288 HOMEBOX (TALE-TYPE).  
 FT DOMAIN 300 303 POLY-ASP.  
 FT DOMAIN 405 418 POLY-GLN.  
 FT DOMAIN 501 516 POLY-GLN.  
 FT DOMAIN 517 528 POLY-HIS.  
 FT DOMAIN 565 572 POLY-SER.  
 FT DOMAIN 613 624 POLY-SER.  
 FT CONFLICT 106 106 C -> R (IN REF. 1).  
 FT CONFLICT 316 316 G -> A (IN REF. 1).  
 FT CONFLICT 678 678 G -> A (IN REF. 1).  
 FT CONFLICT 678 678 G -> A (IN REF. 1).  
 SQ SEQUENCE 693 AA: 73667 MW: FBEEL616493F7EC9 CRC64;  
 Query Match 28.0%; Score 63; DB 1; Length 693;  
 Best Local Similarity 41.2%; Pred. No. 2.6;  
 Matches 14; Conservative 7; Mismatches 9; Indels 4; Gaps 2;  
 QY 7 LCLPTGPGVAIFGCGPVPWFOFTQSMPTLVVY 40  
 DB 141 MALPSGCVGVYGG---PIPSNEQN-PIPSIGV 170  
 AC 067055; STANDARD; PRT; 274 AA.  
 AC 067055; PRT; 274 AA.  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23)  
 DE (Poly(P)/ATP NAD kinase).  
 GN PPNK OR A0\_909.  
 OS Aquifex aeolicus.  
 RT

OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;  
 OC Aquifex.  
 OC NCBI\_TaxID=63363;  
 OX  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VES;  
 RA MEDLINE=98196666; Pubmed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium *Aquifex*  
 RT *aeolicus*.";  
 RT Nature 392:353-358(1998).  
 RL  
 CC -1- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes  
 CC ATP and other nucleoside triphosphates as well as inorganic  
 CC polyphosphate as a source of phosphorus (By similarity).  
 CC -1- CATALYTIC ACTIVITY: ATP + NAD(+) -> ADP + NADP(+).  
 CC -1- COFACTOR: Requires divalent metal ions for activity (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE000713; AAC07026.1; -  
 DR InterPro: IPR002504; ATP\_NADK.  
 DR Pfam: PF01513; NAD\_Kinase; 1.  
 KM Transferrase; Kinase; NAD: NADP: Complete proteome.  
 SQ SEQUENCE 274 AA: 31327 MW: 34510F2B30EB82C6 CRC64;  
 Query Match 27.8%; Score 62.5; DB 1; Length 274;  
 Best Local Similarity 35.1%; Pred. No. 1.2;  
 Matches 13; Conservative 6; Mismatches 13; Indels 5; Gaps 1;  
 QY 6 LCLPTGPGVAIFGCGPVPWFOFTQSMPTLVVY 37  
 DB 166 ILSTPTGTAVALSAGGPIVPSQNLLEFVPCPTL 202  
 AC 095219; STANDARD; PRT; 301 AA.  
 AC 095219; PRT; 301 AA.  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23)  
 DE (Poly(P)/ATP NAD kinase).  
 GN PPNK OR SC01781 OR SC151.21C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetes; Streptomyces; Streptomycetaceae; Streptomyces.  
 OC NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2) / M145;  
 RC MEDLINE=21996410; Pubmed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rebbinowitsch E., Raebdram M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete *Streptomyces*

```

RT coelicolin A3(2).";
RL Nature 417:141-147(2002).
CC -1- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes
CC ATP and other nucleoside triphosphates as well as inorganic
CC polyphosphate as a source of phosphorus (By similarity).
CC -1- CATALYTIC ACTIVITY: ATP + NAD(+) = ADP + NADP(+).
CC -1- COFACTOR: Requires divalent metal ions for activity (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AL109848; CAB52845.1; -.
CC DR InterPro: IPR002504; ATP_NADK.
CC DR Pfam: PF01513; NAD_kinase; 1.
CC KW Transferase: Kinase; NAD; NADP; Complete proteome.
CC SQ SEQUENCE 301 AA; 32081 MW; FCB6F6B01766AFBA CRC64;

Query Match 27.6%; Score 62; DB 1; Length 301;
Best Local Similarity 47.8%; Pred. No. 1.5;
Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 6 LCLPTGPGVAFGGGPPVMPQ 28
DB 184 VLSTPTGTAFAFSAAGGPPVWPE 206

RESULT 4
UP08_ECOLI STANDARD; PRT; 294 AA.
ID UP08_ECOLI
AC P39173; P76913; P76233;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Unknown protein from 2D-page (Spots 726/PR37).
GN YEAD OR B1780.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97251358; PubMed-9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubraman S., Tagami H.,
RA Takada K., Takemoto K., Wada C., Yamamoto Y., Horiiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
RN [3]
RP SEQUENCE OF 1-12.
RN STRAIN-K12 / EMG2;

```

```

RX MEDLINE-97443975; PubMed-9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
RN [4]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE-99420866; PubMed-10493123;
RA Fountoulakis M., Takacs M.-F., Berndt P., Langer H., Takacs B.;
RT "Enrichment of low abundance proteins of Escherichia coli by
RT hydroxyapatite chromatography.";
RL Electrophoresis 20:2181-2195(1999).
CC -1- SIMILARITY: BELONGS TO THE UPF0010 FAMILY. SIGNIFICANTLY RELATED
CC TO MUTAROTASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AE000273; AAC74850.1; ALT_INIT.
CC DR EMBL: D90821; BAA15577.1; -.
CC DR EMBL: D90822; BAA15581.1; -.
CC DR Ecogen: EC12679; yead.
CC DR InterPro: IPR001823; Ald1-epimerase.
CC DR Pfam: PF01263; Aldose-epim; 1.
CC KW Complete proteome.
CC FT ACT_SITE 164
CC FT POTENTIAL. 164
CC SQ SEQUENCE 294 AA; 32666 MW; ADACFE6D46F20957 CRC64;

Query Match 26.4%; Score 59.5; DB 1; Length 294;
Best Local Similarity 38.9%; Pred. No. 3;
Matches 14; Conservative 4; Mismatches 5; Indels 13; Gaps 1;

QY 15 GVAIFGGPVMPWMP-----QFTSMPTYL 37
DB 70 GVAIRGCVPCWPFCAAGCGLPAGCFARNLPWTL 105

RESULT 5
CHAC_ECOLI STANDARD; PRT; 231 AA.
ID CHAC_ECOLI
AC P39163; P77176;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cation transport protein chac.
GN CHAC OR B1218.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NM8191;
RX STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-K12;
RA MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Baba H., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sawai G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horinouchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA, Res. 3:137-155(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EM3L, L28709; AAA20198.1; ALT_INIT.
CC EM3L, AE000220; AAC74302.1; ALT_INIT.
CC EMBL, D90756; BAA36076.1; ALT_INIT.
CC EMBL, D90757; BAA36086.1; ALT_INIT.
CC PccGene, EG12403; chac.
CC Transprot; Complete Proteome.
CC CONFLICT 15 A -> V (IN REF. 1).
CC CONFLICT 79 C -> F (IN REF. 1).
CC FT 25512 MM; 9D708565D53A7C CRC64;
CC SO SEQUENCE 231 AA;
Query Match 26.2%; Score 59; DB 1; Length 231;
Best Local Similarity 47.2%; Pred. No. 2.7;
Matches 17; Conservative 3; Mismatches 12; Indels 4; Gaps 2;
DB 38 LACRPDEGP-VWIFGYGSLMMNPALETESCCTGLV 72
OY 6 LLLCPTGPGVAIFGCGPVMPW---OFTOSMPYTLV 38
ID PPNK_MYCTU STANDARD; PRT; 307 AA.
AC 033196;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23) (Poly(P)/ATP NAD
DE kinase)
GN PPNK OR RV1695 OR MT1734 OR MTC1125.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN-H37RV;
RA MEDLINE=20462915; PubMed=11006082;
RA Kawai S., Mori S., Mukai T., Suzuki S., Yamada T., Hashimoto W.,
RA Murata K.;
RT "Inorganic polyphosphate / ATP-NAD kinase of Mycobacterium flavus and
RT Mycobacterium tuberculosis H37RV.";
RL Biochem. Biophys. Res. Commun. 276:57-63(2000).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

```

```

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmailova M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes
CC ATP and other nucleoside triphosphates as well as inorganic
CC polyphosphate as a source of phosphorus.
CC -1- CATALYTIC ACTIVITY: ATP + NAD(+) = ADP + NADP(+).
CC -1- COFACTOR: Requires divalent metal ions for activity.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AB044336; BAB21478.1; -.
CC EMBL, Z98268; CAB10952.1; -.
CC EMBL, AE007035; AAK46003.1; -.
CC DTGR, MT1734; RV1695; -.
CC DR TuberculList; RV1695; -.
CC DR InterPro; IPR002504; ATP_NADK.
CC DR Pfam; PF01513; NAD_kinase; 1.
CC DR Transference; Kinase; NAD; NADP; Complete proteome.
CC SO SEQUENCE 307 AA; 32903 MM; 68817BE570B6645B CRC64;
Query Match 26.2%; Score 59; DB 1; Length 307;
Best Local Similarity 50.0%; Pred. No. 3.6;
Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;
OY 6 LLLCPTGPGVAIFGCGPVMPW 27
ID 192 LVSTPTGSTAVAFSAGGPVLP 213
DB 192 LVSTPTGSTAVAFSAGGPVLP 213
OY 6 LLLCPTGPGVAIFGCGPVMPW 27
ID PPNK_MYCTU STANDARD; PRT; 311 AA.
AC 049897; 005675;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23) (Poly(P)/ATP NAD
DE kinase)
GN PPNK OR ML1359 OR MLC1351.13C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-TN;
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-TN;
RA MEDLINE=21128732; PubMed=11234002;

```

RA Cole S.T., Eijlmeier K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheelwright P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus.";  
 RL Nature 409:1007-1011(2001).  
 CC -i- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes  
 CC ATP and other nucleoside triphosphates as well as inorganic  
 CC polyphosphate as a source of phosphorus (By similarity).  
 CC -i- CATALYTIC ACTIVITY: ATP + NAD(+) = ADP + NADP(+).  
 CC -i- COFACTOR: Requires divalent metal ions for activity (By  
 CC similarity).  
 CC -i- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -i- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U00021; AAA50923.1; -  
 CC EMBL: 295117; CAB08286.1; -  
 CC EMBL: AL583921; CAC31740.1; -  
 CC Lepidoma; ML1359; -  
 CC DR InterPro: IPR002504; ATP\_NADK.  
 CC Pfam: PF01513; NAD\_kinase; 1.  
 CC Transferrase; Kinase; NAD; NADP; Complete proteome.  
 CC KW SEQUENCE 311 AA; 3328 MW; 0AD2130CC62D3B CRC64;  
 SO  
 Query Match 26.2%; Score 59; DB 1; Length 311;  
 Best Local Similarity 50.0%; Pred. No. 3.7;  
 Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

RA Brouilliet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.M., Connerion I.F., Cummings N.J., Daniel R.A.,  
 RA Denicot F., Devine K.J., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guiseppl G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Paro V., Pohl T.M., Portelle D., Porwolik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Solido B.,  
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takauchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambolt R., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumslein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RL Nature 390:249-256(1997).  
 CC -i- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AB005554; BAA21606.1; -  
 CC EMBL: 299124; CAB16015.1; -  
 CC Subtilisin; Bg11362; yxyc.  
 CC KW Hypothetical protein; Transmembrane; Complete proteome.  
 CC FT TRANSMEM 5 POTENTIAL.  
 CC FT TRANSMEM 137 157 POTENTIAL.  
 CC FT TRANSMEM 183 203 POTENTIAL.  
 CC FT DOMAIN 34 46 POLY-SER.  
 CC FT DOMAIN 57 67 POLY-SER.  
 CC FT DOMAIN 76 81 POLY-SER.  
 CC FT DOMAIN 91 95 POLY-SER.  
 CC SO SEQUENCE 216 AA; 23074 MW; 1B850B1D9EDCA42 CRC64;  
 Query Match 25.8%; Score 58; DB 1; Length 216;  
 Best Local Similarity 47.1%; Pred. No. 3.4;  
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;



[illegible]



RC STRAIN-ATCC 33303 / B10;  
 RX MEDLINE-95362677; PubMed-7635831;  
 RA Pollich M., Kling G.;  
 RT "Identification and sequence analysis of genes involved in late steps  
 in cobalamin (vitamin B12) synthesis in Rhodobacter capsulatus.";  
 RL J. Bacteriol. 177:4481-4487(1995).  
 RN [2].  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SB1003 / St Louis;  
 RX MEDLINE-97404404; PubMed-9256491;  
 RA Vleck C., Paces V., Maltsev N., Paces J., Haselkorn R., Feinstein M.;  
 RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter  
 capsulatus SB1003.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).  
 CC -1- FUNCTION: INVOLVED IN THE CONVERSION OF COBRYIC ACID TO  
 COBANAMIDE. ADDITION OF AMINOPROPANOL ON THE F CARBOXYLIC  
 GROUP (BY SIMILARITY).  
 CC -1- PATHWAY: Cobalamin biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE COBD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z46611; CAA86581.1;  
 DR EMBL: AF010496; AAC16177.1;  
 DR InterPro: IPR004485; Cblb.  
 DR Pfam: PF03186; COBD\_Cblb; 1.  
 DR TIGRfams: TIGR00380; cblb; 1.  
 DR Cobalamin biosynthesis; Transmembrane.  
 KM TRANSMEM 3  
 FT TRANSMEM 57 23  
 FT TRANSMEM 77 77  
 FT TRANSMEM 78 98  
 FT TRANSMEM 154 174  
 FT TRANSMEM 290 310  
 FT TRANSMEM 35 36  
 FT CONFLICT 35 36 GA -> AR (IN REF. 1).  
 SQ SEQUENCE 314 AA; 33407 MW; DE5F599A91DB46D8 CRC64;  
 Query Match 23.6%; Score 53; DB 1; Length 314;  
 Best Local Similarity 57.1%; Pred. No. 20;  
 Matches 12; Conservative 1; Mismatches 6; Indels 2; Gaps 1;  
 QY 7 LCIPRGPGVAIFGGPVPMP 27  
 Db 76 LWDPAWPGVLI-GGILAMP 94  
 RESULT 13  
 CRAR\_HUMAN STANDARD: PRT: 699 AA.  
 ID F48740; O95570; O9UF09;  
 AC F48740; 1996 (Rel. 33, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Complement-activating component of Ra-reactive factor precursor  
 (EC 3.4.21.-) (Ra-reactive factor serine protease p100) (RARF)  
 DE (Mannan-binding lectin serine protease 1) (Mannose-binding protein  
 associated serine protease) (MASP-1).  
 DE GN HAPSI OR CRARF OR CRARF1 OR PRSS5.  
 OS Homo sapiens (Human)  
 OC Buckyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE-94059062; PubMed-8240317;  
 RA Takada F., Takayama Y., Hattuse H., Kawakami M.;

RT "A new member of the C1s family of complement proteins found in a  
 bactericidal factor, Ra-reactive factor, in human serum.";  
 RT Biochem. Biophys. Res. Commun. 196:1003-1009(1993).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Petal liver;  
 RX MEDLINE-94289349; PubMed-8018603;  
 RA Sato T., Endo Y., Matsushita M., Fujita T.;  
 RT "Molecular characterization of a novel serine protease involved in  
 activation of the complement system by mannose-binding protein.";  
 RL Int. Immunol. 6:665-669(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE-97079701; PubMed-8921412;  
 RA Endo Y., Sato T., Matsushita M., Fujita T.;  
 RT "Exon structure of the gene encoding the human mannose-binding  
 protein-associated serine protease light chain: comparison with  
 complement C1r and C1s genes.";  
 RL Int. Immunol. 8:1355-1358(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-9402390; PubMed-10475605;  
 RA Takayama Y., Takada F., Nowatari M., Kawakami M., Matsuura N.;  
 RT "Gene structure of the p100 serine-protease component of the human Ra-  
 reactive factor.";  
 RL Mol. Immunol. 36:505-514(1999).  
 CC -1- FUNCTION: COMPONENT OF THE BACTERICIDAL RA-REACTIVE FACTOR RARF  
 WHICH SPECIFICALLY BINDS TO R1 AND R2 POLYSACCHARIDES EXPRESSED BY  
 CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT  
 CASCADE BY ACTIVATING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE  
 C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.  
 CC -1- SUBUNIT: RARF CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT  
 (CRARF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.  
 CC CRARF IS AN HETERODIMER OF A HEAVY (P70) AND A LIGHT CHAIN (P29)  
 CC LINKED BY A DISULFIDE BOND.  
 CC -1- DOMAIN: CRARF HAS A MODULE ORGANIZATION SIMILAR TO C1R AND C1S.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D17525; BAA04477.1;  
 DR EMBL: D28593; BAA05928.1;  
 DR EMBL: D61695; BAA34864.1;  
 DR EMBL: AB010822; BAA34864.1; JOINED.  
 DR EMBL: AB010813; BAA34864.1; JOINED.  
 DR EMBL: AB010814; BAA34864.1; JOINED.  
 DR EMBL: AB010815; BAA34864.1; JOINED.  
 DR EMBL: AB010816; BAA34864.1; JOINED.  
 DR EMBL: AB010817; BAA34864.1; JOINED.  
 DR EMBL: AB010818; BAA34864.1; JOINED.  
 DR EMBL: AB010819; BAA34864.1; JOINED.  
 DR EMBL: AB010820; BAA34864.1; JOINED.  
 DR EMBL: AB010821; BAA34864.1; JOINED.  
 DR EMBL: D61690; BAA34864.1; JOINED.  
 DR EMBL: D61691; BAA34864.1; JOINED.  
 DR EMBL: D61692; BAA34864.1; JOINED.  
 DR EMBL: D61693; BAA34864.1; JOINED.  
 DR EMBL: D61694; BAA34864.1; JOINED.  
 DR EMBL: AB007617; BAA89206.1;  
 DR EMBL: AB007602; BAA89206.1; JOINED.  
 DR EMBL: AB007603; BAA89206.1; JOINED.  
 DR EMBL: AB007604; BAA89206.1; JOINED.  
 DR EMBL: AB007605; BAA89206.1; JOINED.





OY 12 GCGPVAIF-----GCGPVPMPQFTQSM-----PYTLVYVK 41  
 DB 70 GFGPVPMPFGIPADLPKPEMLLGGGPGPMPMFSDARLHYHPYLGDIK 118

RESULT 15  
 PAC1\_MOUSE STANDARD; PRT; 441 AA.

AC 061644;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Protein kinase C and casein kinase substrate in neurons protein 1.  
 GN PACSIN1 OR PACSIN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
 RX MEDLINE=98417447; PubMed=9746365;  
 RA Plomann M., Lange R., Vopper G., Cramer H., Heinlein U.A.O.,  
 RA Scheif S., Baldwin S.A., Lettges M., Paulsson M.,  
 RA Bartheis D.;  
 RT "PACSIN, a brain protein that is upregulated upon differentiation into  
 RT neuronal cells.";  
 RL Eur. J. Biochem. 256:201-211(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Eye;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP FUNCTION  
 RX PubMed=11082044;  
 RA Modregger J., Rilter B., Witter B., Paulsson M., Plomann M.;  
 RT "All three PACSIN isoforms bind to endocytic proteins and inhibit  
 RT endocytosis.";  
 RL J. Cell Sci. 113:4511-4521(2000).  
 CC -1- FUNCTION: May play a role in vesicle formation and transport.  
 CC -1- SUBUNIT: Homo- and hetero-aggregates with other PACSINs. Binds  
 CC dynamin I, synaptobrevin, synapsin I and the neural Wiskott-Aldrich  
 CC syndrome protein (N-WASP) (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Colocalizes with dynamin I at  
 CC vesicular structures in the cell body and neurites (By  
 CC similarity).  
 CC -1- TISSUE SPECIFICITY: Highly expressed in CNS neurons, especially in  
 CC the pyramidal cells of the hippocampus, Purkinje cells of the  
 CC cerebellum and large neurons of the cortex and brain stem.  
 CC -1- DEVELOPMENTAL STAGE: Expression is seen at embryonic day 17 and is  
 CC upregulated developmentally with a correlation to neuronal  
 CC differentiation.  
 CC -1- PTM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C  
 CC (PKC) (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE PACSIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 FCH DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC  
 CC "This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X85124; CAA59437.1;  
 DR EMBL: BC014698; AAH14698.1;  
 DR HSSP: G60631; IGBO.  
 DR MGD: MGT:1345181; PACSIN1.  
 DR InterPro: IPR001060; Cdc15\_Fes\_CTP4.  
 DR InterPro: IPR001452; SH3.

DR Pfam: PF00611; FCH; 1.  
 DR Pfam: PF00018; SH3; 1.  
 DR PRINTS: PR00452; SH3DOMAIN.  
 DR ProDom: PD000066; SH3; 1.  
 DR SMART: SM00055; FCH; 1.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS50133; FCH; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 KW Endocytosis; SH3 domain; Coiled coil; Phosphorylation.  
 FT DOMAIN 10 73 FCH.  
 FT DOMAIN 382 441 SH3.  
 FT DOMAIN 144 165 COILED COIL (POTENTIAL).  
 FT DOMAIN 183 217 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 441 AA; 50575 MW; 21BBE339A14A41F9 CRC64;

Query Match 23.3%; Score 52.5; DB 1; Length 441;  
 Best Local Similarity 37.9%; Pred. No. 32;  
 Matches 11; Conservative 5; Mismatches 4; Indels 9; Gaps 2;

OY 11 TGCGVAIFGCGPVPMPQFTQ---SMPT 36  
 DB 287 TSGPGM-----PMNMPQFEEHNPDLPH 309

Search completed: May 20, 2003, 16:35:47  
 Job time : 12.0385 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:19:09 ; Search time 50.6868 Seconds  
(without alignments)  
166.669 Million cell updates/sec

Title: US-09-869-155-19  
Perfect score: 225  
Sequence: 1 VANRFLCLPTGPGVAIFG.....GVPMPQFTQSMPTLVVVK 41

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMHL\_21:\*  
2: sp-archaea:\*  
3: sp-bacteria:\*  
4: sp-fungi:\*  
5: sp-human:\*  
6: sp-invertebrate:\*  
7: sp-mammal:\*  
8: sp-organeller:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-virus:\*  
13: sp-vertebrate:\*  
14: sp-unclassified:\*  
15: sp-virus:\*  
16: sp-bacteriap:\*  
17: sp-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111	49.3	419	10	08S1V0
2	96	42.7	424	10	08S1V1
3	79	35.1	422	10	08S1V3
4	77.5	34.4	434	10	08S1V5
5	71	31.6	402	10	08S1U9
6	68	30.2	492	16	08XT86
7	67.5	30.0	434	10	08S1U4
8	61	27.1	362	2	08V0U9
9	61	27.1	446	16	09A7W1
10	59.5	26.4	301	16	08XD06
11	59.5	26.4	433	10	09ZVS4
12	59.5	26.4	534	17	097XY4
13	59	26.2	434	10	09ZVS5
14	57.5	25.6	568	11	08VH03
15	57	25.3	263	4	09NV23
16	57	25.3	351	11	08R3G1

17	56.5	25.1	283	16	08RAC3	08rac3 thermomane
18	56.5	25.1	1054	5	016331	016331 caenorhabd
19	56	24.9	291	16	08REK9	08rek9 yersinia pe
20	56	24.9	305	2	053129	053129 rhodobacter
21	56	24.9	650	8	09XPH4	09xph4 beta vulgar
22	55.5	24.7	750	11	09CW61	09cw61 mus musculu
23	55.5	24.7	1050	5	016436	016436 caenorhabd
24	55	24.4	365	12	08ORT0	08ort0 chimpanzee
25	55	24.4	629	2	007639	007639 streptomyce
26	54.5	24.2	193	11	099ML1	099ml1 mus musculu
27	54.5	24.2	328	5	095TD3	095td3 caenorhabd
28	54	24.0	195	16	08XSR5	08xsr5 raiistonla s
29	54	24.0	232	4	09H2W2	09h2w2 homo sapien
30	54	24.0	288	16	09PB74	09pb74 xyella fas
31	54	24.0	331	2	087941	087941 thauera aro
32	54	24.0	546	4	096M09	096m09 homo sapien
33	54	24.0	559	4	09H5P1	09h5p1 homo sapien
34	54	24.0	621	10	09LGX7	09lgx7 oryza sativ
35	54	24.0	1302	4	09C092	09c092 homo sapien
36	54	24.0	1980	5	09NDY5	09ndy5 leishmania
37	54	24.0	2552	2	09XCF2	09xcf2 mycobacteri
38	53.5	23.8	325	16	09PD03	09pd03 xyella fas
39	53.5	23.8	463	10	039496	039496 cylindrothe
40	53.5	23.8	671	5	09VB83	09vb83 drosophila
41	53.5	23.8	684	5	024035	024035 drosophila
42	53.5	23.8	768	16	086766	086766 streptomyce
43	53.5	23.8	831	5	08TA47	08ta47 drosophila
44	53	23.6	293	11	09D6L4	09d6l4 mus musculu
45	53	23.6	294	2	09R6G6	09r6g6 agrobacteri

#### ALIGNMENTS

RESULT 1  
ID 08S1V0 PRELIMINARY; PRT; 419 AA.  
AC 08S1V0;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Putative dermal glycoprotein.  
GN P0504E02.9.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1. PAC  
clone:P0504E02.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AP003269; BAB89708.1; -  
SQ SEQUENCE 419 AA; 43786 MW; 792B55FAAF3F3A8 CRC64;  
Query Match 49.3%; Score 111; DB 10; Length 419;  
Best Local Similarity 55.3%; Pred. No. 7.7e-06;  
Matches 26; Conservative 4; Mismatches 11; Indels 6; Gaps 2;  
QY 1 VANRFLCLPT--GGPGVAIFGGGPV---PMPQFTQSMPTLVVVK 41  
DB 195 VAGRFLLCLPRLGYGGVAIFGGGPVYIGSLPDTFTTLDYPLVAK 241  
RESULT 2  
ID 08S1V1 PRELIMINARY; PRT; 424 AA.  
AC 08S1V1;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

```
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative dermal glycoprotein.
GN P0504E02.8
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponebare(GA3) genomic DNA, chromosome 1, PAC
clone:P0504E02."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003269; BAB89707.1; -
SQ SEQUENCE 424 AA; 44620 MW; 0DADC5CAC8852C34 CRC64;

Query Match 42.7%; Score 96; DB 10; Length 424;
Best Local Similarity 73.1%; Pred. No. 0.00057;
Matches 19; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 1 VANFLLCLPT--GGPGVAIFGGGPV 24
DB 194 VANFALCLPSSGGGDSVAIFGGPL 219
|||||
ID 08SIU9 PRELIMINARY; PRT; 422 AA.
AC 08SIU9;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE Putative dermal glycoprotein.
GN P0504E02.6.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponebare(GA3) genomic DNA, chromosome 1, PAC
clone:P0504E02."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003269; BAB89705.1; -
SQ SEQUENCE 422 AA; 44745 MW; 08CFF26346400A9F CRC64;

Query Match 35.1%; Score 79; DB 10; Length 422;
Best Local Similarity 45.2%; Pred. No. 0.072;
Matches 19; Conservative 5; Mismatches 8; Indels 10; Gaps 3;

QY 1 VANFLLCLPTGGPGVAIFGGGPV---VPW--PQFTQSMPT 36
DB 196 VAKFALCLPS---VAVFGGPFVLIFPYSRPDIMOKLSVT 233
|||||
ID 08SIU5 PRELIMINARY; PRT; 434 AA.
AC 08SIU5;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE Putative dermal glycoprotein.
GN P0504E02.3.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
```

```
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponebare(GA3) genomic DNA, chromosome 1, PAC
clone:P0504E02."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003269; BAB89703.1; -
SQ SEQUENCE 434 AA; 45342 MW; 7E2549288A6B28B1 CRC64;

Query Match 34.4%; Score 77.5; DB 10; Length 434;
Best Local Similarity 66.7%; Pred. No. 0.11;
Matches 16; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 VANFLLCLPT--TGSGPGVAIFGGGP 23
DB 197 VERFALCLPGGGGTGAIFGGP 220
|||||
ID 08SIU9 PRELIMINARY; PRT; 402 AA.
AC 08SIU9;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE Putative dermal glycoprotein.
GN P0504E02.10.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponebare(GA3) genomic DNA, chromosome 1, PAC
clone:P0504E02."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003269; BAB89709.1; -
SQ SEQUENCE 402 AA; 41793 MW; E8C64A214D6F6720 CRC64;

Query Match 31.6%; Score 71; DB 10; Length 402;
Best Local Similarity 47.2%; Pred. No. 0.68;
Matches 17; Conservative 3; Mismatches 8; Indels 8; Gaps 2;

QY 4 RFLCLPT--GGPGVAIFGGGPVWPQFTQSMPT 37
DB 181 QFAVCLPSTAAAPGVAIFGGGPV-----NLMPTL 210
|||||
ID 08XT86 PRELIMINARY; PRT; 492 AA.
AC 08XT86;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE Putative benzaldehyde dehydrogenase oxidoreductase protein
(BC 1.2.1.28).
GN RSP0229 OR RS05194.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=2161879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
```

RA Ariat M., Billault A., Brotlier P., Camus J.C., Catolico L.,  
 RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurio W., Schlex T.,  
 RA Guiguer P., Thibault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.,  
 RT Genome sequence of the plant pathogen *Ralstonia solanacearum*.  
 RL Nature 415:497-502(2002).  
 DR EMBL: AL646077; CAD17380.1;  
 DR InterPro: IPR002086; Aldehyde\_dehydr.  
 DR Pfam: PF00171; aldehyd. 1.  
 KM Oxidoreductase, plasmid, complete proteome.  
 SQ SEQUENCE 492 AA; 52058 MW; F795296E3EF7EC6 CRC64;

Query Match 30.2%; Score 68; DB 16; Length 492;  
 Best Local Similarity 45.7%; Pred. No. 1.9;  
 Matches 16; Conservative 4; Mismatches 9; Indels 6; Gaps 2;

QY 10 PTGGPGVAIFG---GGPVPWPOFTQSMPTLVYVK 41  
 DB 453 PFGPGIAGNGTSMGPGADWEYTO---WRWVTVK 484

## RESULT 7

ID 08S104 PRELIMINARY; PRT; 434 AA.  
 AC 08S104;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE P0504E02.16 protein.  
 GN P0504E02.16.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriatloideae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GAS) genomic DNA, chromosome 1, PAC  
 clone:P0504E02.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003269; BAB89714.1;  
 SQ SEQUENCE 434 AA; 46236 MW; 87C36211DD57F72 CRC64;

Query Match 30.0%; Score 67.5; DB 10; Length 434;  
 Best Local Similarity 41.5%; Pred. No. 2;  
 Matches 17; Conservative 4; Mismatches 9; Indels 11; Gaps 2;

QY 3 NRFLCLPTGGPGVAIFGGPV-----PWPOFTQSMPT 36  
 DB 200 NKEALCL----PGFAFGDPVYIGTESLGIVNTESLPYT 236

## RESULT 8

ID 08VUL9 PRELIMINARY; PRT; 362 AA.  
 AC 08VUL9;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Poly(P)/ATP-NAD kinase.  
 GN MNK.  
 OS Micrococcus luteus (Micrococcus lysodeikticus).  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Micrococcales; Micrococcaceae; Micrococcus.  
 OC NCBI\_TaxID=1270;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IFO 3242;  
 RA Kawai S., Murata K.;  
 RT "Direct phosphorylation of NADH by NAD kinase.";

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB070351; BAB84189.1;  
 DR InterPro: IPR002504; ATP\_NADK.  
 DR Pfam: PF01513; NAD\_kinase. 1.  
 KW Kinase.  
 SQ SEQUENCE 362 AA; 39114 MW; 106EC546E58F47E1 CRC64;

Query Match 27.1%; Score 61; DB 2; Length 362;  
 Best Local Similarity 50.0%; Pred. No. 11;  
 Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 6 LLCLPTGGPGVAIFGGPVPW 27  
 DB 188 VLAPPTGSTAYAFSGCPVWVP 209

## RESULT 9

ID 09A7M1 PRELIMINARY; PRT; 446 AA.  
 AC 09A7M1;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Succinylarginine dihydrolase.  
 GN CC1608.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 OC NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A.S., Stephens C., Phadke N.D., Ely B.,  
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Uitterback T., Tran K., Wolf A., Yamthayan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of *Caulobacter crescentus*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL: AE005836; AAK23587.1;  
 DR TIGR: CC1608;  
 KW Hydrolase; Complete proteome.  
 SQ SEQUENCE 446 AA; 48314 MW; 406C09EDC6AF7B9C CRC64;

Query Match 27.1%; Score 61; DB 16; Length 446;  
 Best Local Similarity 39.4%; Pred. No. 13;  
 Matches 13; Conservative 2; Mismatches 18; Indels 0; Gaps 0;

QY 2 ANRFLCLPTGGPGVAIFGGPVPWPOFTQSMPT 34  
 DB 178 ANHYRLCAEHGPGVNLFWGREAMSHWDGRFP 210

## RESULT 10

ID 08XD06 PRELIMINARY; PRT; 301 AA.  
 AC 08XD06;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Orf, hypothetical protein.  
 GN YEAD OR 22820 OR ECS2489.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;

RA Parra N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobbeck E.J., Davis N.W., Lim A., Dimlantia E.T., Potamoukis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.,  
 RT "genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533(2001).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-0157:H7 / RIMD 0509957;  
 RX MEDLINE-21156231; PubMed-11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
 RA Kubara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL: AE005401; AAC56769.1; -;  
 DR EMBL: AP002558; BAB35912.1; -;  
 DR InterPro: IPR001823; A1d1.epimerase.  
 DR Pfam: PF01263; Aldose-epim; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 301 AA; 33590 MW; 7F56DA678B6DE326 CRC64;  
 Query Match 26.4%; Score 59.5; DB 16; Length 301;  
 Best Local Similarity 38.9%; Pred. No. 14;  
 Matches 14; Conservative 4; Mismatches 5; Indels 13; Gaps 1;  
 QY 15 GVAIFGCGPVPW-----OFTGSMPTL 37  
 Db 77 GVAIRGVPWCPWFPGPAACGLPAHGFARMLPTL 112  
 ID 092VVS4 PRELIMINARY; PRT; 433 AA.  
 AC 092VVS4;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE F15K9.17 protein (Hypothetical 45.7 kDa protein) (Putative  
 DE extracellular dermal glycoprotein EDGP).  
 GN F15K9.17.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Kymenetskaia V.S., Schwartz J.R., Toriumi M., Yu G., Li J., Liu S.,  
 RA Jember K., Feng J., Kim C., Li Y., Shinn P., Sun H., Davis R.W.,  
 RA Ecker J.R., Federspiel N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F15K9 sequence.";  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Theologis A.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Theologis A.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 DR EMBL: AE006771; AK41784.1; -;  
 KW Complete proteome.

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
 RA Palm C.J., Theologis A., Ecker J., Davis R.W.,  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [16]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,  
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,  
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,  
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,  
 RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,  
 RA Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene F15K9.17 (GI3850579).";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [17]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,  
 RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,  
 RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,  
 RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,  
 RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,  
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene F15K9.17 (GI3850579).";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC005278; AAC72119.1; -;  
 DR EMBL: AF325092; AAK17160.1; -;  
 DR EMBL: AY035026; AAL15204.1; -;  
 DR EMBL: AY035026; AAK59531.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 433 AA; 45717 MW; 7214FC4B8BA72962 CRC64;  
 Query Match 26.4%; Score 59.5; DB 10; Length 433;  
 Best Local Similarity 60.0%; Pred. No. 19;  
 Matches 12; Conservative 2; Mismatches 5; Indels 1; Gaps 1;  
 QY 4 RFLCLPFGCGVAFPGGP 23  
 Db 199 KFAVCL-TSGKGVAFPGGP 217  
 ID 097XY4 PRELIMINARY; PRT; 534 AA.  
 AC 097XY4;  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Nitric oxide reductase, subunit I (cytochrome B) (norB-1).  
 GN NORB-1 OR SS01571.  
 OS Sulfolobus solfataricus.  
 OS Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OC NCBI\_TaxID=2287;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE-21332296; PubMed-11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aways M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moers A., Erasuo G., Fletcher C., Gordon P.M.K.,  
 RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.L., Doollittle W.F., Duguet M., Gasterland T.,  
 RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 DR EMBL: AE006771; AK41784.1; -;  
 KW Complete proteome.

SEQ SEQUENCE 534 AA: 58445 MW: 4E9C344AF3984962 CRC64:

Query Match 26.4%: Score 59.5; DB 17; Length 534;

Best Local Similarity 38.7%: Pred. No. 24; Mismatches 12; Conservative 7; Mismatches 9; Indels 3; Gaps 2;

QY 13 GPGV-AIFGGGPV--PWPQFTQSMPTLVV 40

DB 240 GSGIGAFSGSLPIAPWPNTEDQFLWIMI 270

RESULT 13

Q9ZVS5

ID 09ZVS5 PRELIMINARY: PRT: 434 AA.

AC 09ZVS5:

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DE F15K9.16 protein (Putative extracellular dermal glycoprotein precursor).

GN F15K9.16.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Vysotskaya V.S., Schwartz J.R., Toriumi M., Yu G., Li J., Liu S.,

RA Kremenetska I., Juro J., Araujo R., Buehler E., Conway A.B.,

RA Dwyer K., Feng C., Kim C., Li Y., Shinn P., Sun H., Davis R.W.,

RA Ecker J.R., Federspiel N.A., Theologis A.;

RT "Arabidopsis thaliana chromosome 1 BAC F15K9 sequence."

RL Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Theologis A.;

RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Theologis A.;

RL Submitted (Nov-1998) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Theologis A.;

RL Submitted (Nov-1998) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Theologis A.;

RL Submitted (Nov-1998) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Theologis A.;

RL Submitted (Nov-1998) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Theologis A.;

RL Submitted (Nov-1998) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Theologis A.;

RL Submitted (Nov-1998) to the EMBL/GenBank/DBJ databases.

ID 08VHQ3 PRELIMINARY: PRT: 568 AA.

AC 08VHQ3:

DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Protein phosphatase 1 regulatory subunit 16B.

GN ppp1r16B.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/C; TISSUE=CEREBELLUM;

RA Northcutt G.M., Kurschner C.;

RT "A novel mouse protein phosphatase 1 targeting subunit is expressed in three distinct regions of the brain."

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DE EMBL: AF423761; AAL62093.1;

DR InterPro: IPR002110; ANK.

DR InterPro: IPR001230; Prenyl\_site.

DR Pfam: PF00023; ank.5.

DR PRINTS: PR01415; ANKYRIN.

DR SMART: SM00248; ANK.5.

DR PROSITE: PS00088; ANK\_REPEAT.4.

DR PROSITE: PS00297; ANK\_REPEAT\_REGION.1.

DR PROSITE: PS00294; PRENYLATION; UNKNOWN.1.

SEQUENCE 568 AA: 63571 MW: 5D8446C81DEBA3D1 CRC64:

Query Match 25.6%: Score 57.5; DB 11; Length 568;

Best Local Similarity 43.3%: Pred. No. 45; Mismatches 13; Conservative 3; Mismatches 11; Indels 3; Gaps 1;

QY 12 GPGVAFGGGPVWPQFTQSMPTLVV 41

DB 459 GNPQVA---DVPPWSGFQSPQTLLELK 485

RESULT 15

Q9NVZ3

ID 09NVZ3 PRELIMINARY: PRT: 263 AA.

AC 09NVZ3:

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE CDNA FLJ10420 f1s, clone NT2RP1000170 (Hypothetical 28.3 kDa protein).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN

RP SEQUENCE FROM N.A.

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,

RA Magatsuna M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,

RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,

RA Yamamoto S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,

RA Watanabe J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,

RA Niinomiya K., Iwayanagi T.;

RT "NEO human cDNA sequencing project."

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC STRAIN=COLON;

RA Strausberg R.;

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC STRAIN=RETINA;

RA Strausberg R.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DE EMBL: AK001282; BAA91598.1;

RESULT 14

Q8VHQ3





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:35:02 ; Search time 18.9231 seconds  
(without alignments)  
63.750 Million cell updates/sec

Title: US-09-869-155-19  
Perfect score: 225  
Sequence: 1 VANRFLCLPTGPGVAIFG.....GPVWPQFTQSPYTLVVK 41

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2-6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2-6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2-6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2-6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2-6/ptodata/1/1aa/PGTUS.COMB.pep.\*  
6: /cgn2-6/ptodata/1/1aa/Backfilltest.pep.\*  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.5	24.2	504	US-09-126-420A-18	Sequence 18, Appl
2	53	23.6	251	US-08-944-483-47	Sequence 47, Appl
3	53	23.6	304	US-09-088-651-2	Sequence 2, Appl
4	52.5	23.3	441	US-08-630-915A-34	Sequence 34, Appl
5	52.5	23.3	635	US-09-081-975-3	Sequence 3, Appl
6	51	22.7	45	US-08-881-450A-16	Sequence 16, Appl
7	51	22.7	452	US-08-317-880-2	Sequence 2, Appl
8	51	22.7	452	US-08-782-396-2	Sequence 2, Appl
9	50.5	22.4	504	US-09-126-420A-23	Sequence 23, Appl
10	50	22.2	218	US-09-081-975-7	Sequence 7, Appl
11	50	22.2	281	US-08-487-748A-9	Sequence 9, Appl
12	50	22.2	281	US-08-487-748A-10	Sequence 10, Appl
13	50	22.2	281	US-08-480-070C-10	Sequence 10, Appl
14	50	22.2	281	US-08-829-525-10	Sequence 10, Appl
15	50	22.2	281	US-08-609-583A-10	Sequence 10, Appl
16	50	22.2	281	US-08-937-399-10	Sequence 10, Appl
17	50	22.2	281	US-09-310-367-10	Sequence 10, Appl
18	50	22.2	281	US-09-032-337-10	Sequence 10, Appl
19	49.5	22.0	433	US-09-046-158A-2	Sequence 2, Appl
20	49.5	22.0	897	US-07-960-389-2	Sequence 2, Appl
21	49	21.8	1266	US-08-506-296B-4	Sequence 4, Appl
22	48.5	21.6	543	US-09-535-008-63	Sequence 63, Appl
23	48.5	21.6	577	US-08-996-138-15	Sequence 15, Appl
24	48.5	21.6	625	US-08-995-659-15	Sequence 15, Appl
25	48.5	21.6	625	US-09-215-649A-15	Sequence 15, Appl
26	48.5	21.6	625	US-09-577-780-15	Sequence 15, Appl

28	48.5	21.6	1088	US-09-082-059-2	Sequence 2, Appl
29	48.5	21.6	1647	US-09-535-008-67	Sequence 67, Appl
30	48.5	21.6	1646	US-09-535-008-2	Sequence 2, Appl
31	48.5	21.6	1649	US-09-535-008-75	Sequence 75, Appl
32	48.5	21.6	1650	US-09-535-008-71	Sequence 71, Appl
33	48.5	21.6	1678	US-09-535-008-69	Sequence 69, Appl
34	48.5	21.6	1679	US-09-535-008-65	Sequence 65, Appl
35	48.5	21.6	1681	US-09-535-008-77	Sequence 77, Appl
36	48.5	21.6	1682	US-09-535-008-73	Sequence 73, Appl
37	48	21.3	221	US-09-134-001C-3253	Sequence 3253, Ap
38	48	21.3	525	US-08-336-340-2	Sequence 2, Appl
39	48	21.3	525	US-08-786-555-2	Sequence 2, Appl
40	48	21.3	1105	US-08-999-774A-2	Sequence 2, Appl
41	48	21.3	2512	US-08-801-263A-9	Sequence 9, Appl
42	48	21.3	2512	US-09-102-248-9	Sequence 9, Appl
43	47.5	21.1	405	US-09-413-574-2	Sequence 2, Appl
44	47.5	21.1	614	US-09-832-436-2	Sequence 2, Appl
45	47.5	21.1	614	US-09-832-614A-2	Sequence 2, Appl

## ALIGNMENTS

```
RESULT 1
US-09-126-420A-18
; Sequence 18, Application US/09126420A
; Patent No. 6376753
; GENERAL INFORMATION:
; APPLICANT: BATARD, YANNICK
; APPLICANT: ROBINEAU, TIBURCE
; APPLICANT: DURST, FRANCIS
; APPLICANT: WERCK-REICHAUT, DANIELE
; APPLICANT: DIDIERJEAN, LUC
; TITLE OF INVENTION: PURIFIED CYTOCHROME P450 CYP76B1 FROM HELIANTHUS
; TITLE OF INVENTION: TUBEROSUS AND ITS APPLICATIONS AS BIOCATALYST IN
; TITLE OF INVENTION: PARTICULAR FOR THE DEGRADATION OF ENVIRONMENTAL
; FILE REFERENCE: 03715.0032
; CURRENT APPLICATION NUMBER: US/09/126,420A
; CURRENT FILING DATE: 1996-07-30
; PRIOR APPLICATION NUMBER: 60/054,351
; PRIOR FILING DATE: 1997-07-31
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Solanum melongena
; US-09-126-420A-18

Query Match      24.2%  Score 54.5;  DB 4;  Length 504;
Best Local Similarity 30.6%;  Pred. No. 16;
Matches 15;  Conservative 3;  Mismatches 10;  Indels 21;  Gaps 2;

OY 10 PPGGPGVAIFGCG-----GPVWPQFTQSPYTLV 39
DB 35 PPGPGPLIFGKMFELGTPEYKKMNVLRKYGCVLM--LKEGSTYIMV 81

RESULT 2
US-08-944-483-47
; Sequence 47, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLAAS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
```

TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
TITLE OF INVENTION: OF THE PROSTATE  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,483  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6183.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 251 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: NO. 62324566  
US-08-944-483-47

Query Match 23.6%; Score 53; DB 4; Length 251;  
Best Local Similarity 36.2%; Pred. No. 12;  
Matches 17; Conservative 6; Mismatches 14; Indels 10; Gaps 3;

Qy 1 VANRFL-LCLPTG---GPGVAIFGGGVPWPQFTQSMPTLVVYK 41  
Db 115 VNAAFVMPICLPBGPGQEGAMVYSGW---KQFLQRPETLMEIE 157

RESULT 3  
US-09-088-651-2  
Sequence 2, Application US/09088651  
Patent No. 6165771  
GENERAL INFORMATION:  
APPLICANT: BURGESS, NICOLA A.  
APPLICANT: CLINKENBEARD, HELEN E.  
APPLICANT: SOUTHAN, CHRISTOPHER D.  
TITLE OF INVENTION: NOVEL COMPOUNDS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RATNER & PRESTIA  
STREET: P.O. BOX 980  
CITY: VALLEY FORGE  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/088,651  
FILING DATE: JUNE 1, 1998

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB9712088.5  
FILING DATE: 10-JUNE-1997  
APPLICATION NUMBER: EP 97308295.1  
FILING DATE: 17-OCT-1997  
APPLICATION NUMBER: GB 9803650.2  
FILING DATE: 20-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F.  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH30358  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 304 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-088-651-2

Query Match 23.6%; Score 53; DB 4; Length 304;  
Best Local Similarity 45.2%; Pred. No. 14;  
Matches 14; Conservative 2; Mismatches 13; Indels 2; Gaps 2;

Qy 10 PTGGGVAIFGGGVPWPQFTQSMPTLVVY 40  
Db 241 PT-GPGCG-WGLGVPVPMVIMQTEPSPVLIRV 269

RESULT 4  
US-08-630-915A-34  
Sequence 34, Application US/08630915A  
Patent No. 6309820  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: HOFFMAN, NO. 6309820h  
APPLICANT: KAY, Brian K.  
APPLICANT: FOMLAKES, Dana M.  
APPLICANT: MCCONNELL, Stephen J.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
TITLE OF INVENTION: USING SAME  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 34:



APPLICATION NUMBER: US/08/317,880  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rogalsky, Peter  
REGISTRATION NUMBER: 38,601  
REFERENCE/DOCKET NUMBER: 19226/610 (R-5261)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1634  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 452 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-880-2

Query Match 22.7% Score 51; DB 1; Length 452;  
Best Local Similarity 50.0%; Pred. No. 41;  
Matches 14; Conservative 2; Mismatches 8; Indels 4; Gaps 2;

OY 9 LPTGGPGV--AIFGGGPV--PWPQFTOS 32  
||| ||| : ||| ||| ||| |||  
Db. 360 LPNGPGMTASISOGSGVTPGPYIAOS 387

RESULT 8  
US-08-782-396-2  
Sequence 2, Application US/08782396  
Patent No. 5871940

GENERAL INFORMATION:  
APPLICANT: Hall, Linda M.  
APPLICANT: Feng, Guoping  
TITLE OF INVENTION: PROTEINS FOR CATION CHANNEL EXPRESSION  
TITLE OF INVENTION: AND FUNCTION  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/782,396  
FILING DATE: 13-JAN-1997  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/317,880  
FILING DATE: 04-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Rogalsky, Peter  
REGISTRATION NUMBER: 38,601  
REFERENCE/DOCKET NUMBER: 19226/611 (R-5261B)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1634  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 452 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-782-396-2

Query Match 22.7% Score 51; DB 2; Length 452;  
Best Local Similarity 50.0%; Pred. No. 41;  
Matches 14; Conservative 2; Mismatches 8; Indels 4; Gaps 2;

OY 9 LPTGGPGV--AIFGGGPV--PWPQFTOS 32  
||| ||| : ||| ||| ||| |||  
Db. 360 LPNGPGMTASISOGSGVTPGPYIAOS 387

RESULT 9  
US-09-126-420A-23

Sequence 23, Application US/09126420A  
Patent No. 6376753  
GENERAL INFORMATION:  
APPLICANT: BARTARD, YANNICK  
APPLICANT: ROBINEAU, TIBURCE  
APPLICANT: DURST, FRANCIS  
APPLICANT: WERCK-REICHART, DANIELE  
APPLICANT: DIDIERJEAN, LUC  
TITLE OF INVENTION: PURIFIED CYTOCHROME P450 CYP76B1 FROM HELIANTHUS  
TITLE OF INVENTION: PARTICULAR FOR THE DEGRADATION OF ENVIRONMENTAL  
TITLE OF INVENTION: POLLUTANTS AND FOR ALTERING THE RESISTANCE OF PLANTS S  
FILE REFERENCE: 03715.0032  
CURRENT APPLICATION NUMBER: US/09/126,420A  
CURRENT FILING DATE: 1998-07-30  
PRIOR FILING DATE: 1997-07-31  
PRIOR FILING DATE: 1997-07-31  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 23  
LENGTH: 504  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-126-420A-23

Query Match 22.4% Score 50.5; DB 4; Length 504;  
Best Local Similarity 43.5%; Pred. No. 54;  
Matches 10; Conservative 2; Mismatches 4; Indels 7; Gaps 1;

OY 22 GPVWPQPF-----TOSMPYTL 37  
||| ||| : ||| ||| ||| |||  
Db. 39 GPVWPVPLGNLLQVLDLNMPSYL 61

RESULT 10  
US-09-081-975-7

Sequence 7, Application US/09081975  
Patent No. 6451979  
GENERAL INFORMATION:  
APPLICANT: Kaelin, William  
APPLICANT: Jost, Christine  
TITLE OF INVENTION: METHODS OF TREATMENT USING  
TITLE OF INVENTION: NBS-1, ANTIBODIES AND PROTEINS THEREO, AND USES OF THE  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon Peabody LLP  
STREET: 101 Federal Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/081,975  
FILING DATE: 12-MAY-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

QY 3 NRFLC--LPTGGPVAI-FGGGPVWPQTQSM 33

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,070C  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-024  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 281 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-480-070C-10

Query Match 22.2% Score 50; DB 3; Length 281;  
Best Local Similarity 40.0%; Pred. No. 33;  
Matches 14; Conservative 4; Mismatches 13; Indels 4; Gaps 2;

QY 3 NRFLC---LPTGGPGVAI-FGGGPPWPQFTQSM 33  
| : | | | | | : | : | | | | :  
Db 33 NAYLPCSYTLPTSGTLVPMCMGKFCFWSQCTNEL 67

RESULT 14  
US-08-829-525-10  
Sequence 10, Application US/08829525  
Patent No. 6084083  
GENERAL INFORMATION:  
APPLICANT: Levinson, Douglas A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
TREATMENT AND DIAGNOSIS OF IMMUNE DISORDERS  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036/2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/829,525  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/609,583  
FILING DATE: 01-MAR-1996  
APPLICATION NUMBER: US 08/487,748  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/398,633  
FILING DATE: 03-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-081  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:

LENGTH: 281 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-829-525-10

Query Match 22.2% Score 50; DB 3; Length 281;  
Best Local Similarity 40.0%; Pred. No. 33;  
Matches 14; Conservative 4; Mismatches 13; Indels 4; Gaps 2;

QY 3 NRFLC---LPTGGPGVAI-FGGGPPWPQFTQSM 33  
| : | | | | | : | : | | | | :  
Db 33 NAYLPCSYTLPTSGTLVPMCMGKFCFWSQCTNEL 67

RESULT 15  
US-08-609-583A-10  
Sequence 10, Application US/08609583A  
Patent No. 6204371  
GENERAL INFORMATION:  
APPLICANT: Levinson, Douglas A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
TREATMENT AND DIAGNOSIS OF IMMUNE DISORDERS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036/2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/609,583A  
FILING DATE: 01-MAR-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,748  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/398,633  
FILING DATE: 03-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-048  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 281 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-609-583A-10

Query Match 22.2% Score 50; DB 4; Length 281;  
Best Local Similarity 40.0%; Pred. No. 33;  
Matches 14; Conservative 4; Mismatches 13; Indels 4; Gaps 2;

QY 3 NRFLC---LPTGGPGVAI-FGGGPPWPQFTQSM 33  
| : | | | | | : | : | | | | :  
Db 33 NAYLPCSYTLPTSGTLVPMCMGKFCFWSQCTNEL 67

Search completed: May 20, 2003, 16:44:46  
Job time : 20.9231 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:49:49 ; Search time 20.2747 Seconds  
(without alignments)  
200.549 Million cell updates/sec

Title: US-09-869-155-19

Perfect score: 225  
Sequence: 1 VANRFLCPTGPGVAIFG.....GPVMPQFTQSMPTLVVK 41

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications -AA:\*  
1: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the entry being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	27.6	291	9	US-09-738-626-5062
2	54	24.0	1220	9	US-09-736-968A-14
3	54	24.0	1220	10	US-09-736-969A-14
4	54	24.0	1220	10	US-09-736-969A-14
5	54	24.0	2050	9	US-09-736-968A-107
6	54	24.0	2050	10	US-09-736-969A-93
7	54	24.0	2050	10	US-09-736-969A-93
8	53.5	23.8	73	10	US-09-867-550-1298
9	53.5	23.8	91	10	US-09-864-761-41973
10	53.5	23.8	684	10	US-09-823-240-9
11	53	23.6	81	9	US-10-023-282-740
12	53	23.6	114	9	US-10-023-282-739
13	53	23.6	136	9	US-09-764-891-4761
14	53	23.6	235	9	US-10-023-282-318
15	53	23.6	235	9	US-10-023-282-738
16	53	23.6	361	9	US-10-149-819-10
17	53	23.6	679	10	US-09-874-198-6
18	53	23.6	679	10	US-09-874-238-6
19	52.5	23.3	441	10	US-09-879-957-34

20	52.5	23.3	519	9	US-10-029-180-8	Sequence 8, Appl
21	52.5	23.3	600	9	US-10-029-180-120	Sequence 120, App
22	52.5	23.3	611	9	US-10-029-180-119	Sequence 119, App
23	52.5	23.3	635	12	US-10-155-059-3	Sequence 3, Appl
24	51	22.7	122	9	US-10-102-806-522	Sequence 522, App
25	50.5	22.4	136	10	US-09-948-018-6	Sequence 6, Appl
26	50.5	22.4	355	10	US-09-948-018-2	Sequence 2, Appl
27	50.5	22.4	380	10	US-09-948-018-36	Sequence 36, Appl
28	50	22.2	110	10	US-09-864-761-40789	Sequence 40789, A
29	50	22.2	218	9	US-10-097-065-230	Sequence 230, App
30	50	22.2	218	12	US-10-155-059-7	Sequence 7, Appl
31	50	22.2	281	9	US-10-004-633-10	Sequence 10, Appl
32	50	22.2	414	9	US-10-125-635A-406	Sequence 406, App
33	50	22.2	414	9	US-09-938-864-406	Sequence 406, App
34	50	22.2	636	10	US-09-732-384-10	Sequence 10, Appl
35	49.5	22.0	245	9	US-10-153-668-92	Sequence 92, Appl
36	49.5	22.0	897	9	US-10-099-895-1	Sequence 1, Appl
37	48.5	21.6	110	10	US-09-732-665-12	Sequence 12, Appl
38	48.5	21.6	152	10	US-09-732-665-11	Sequence 11, Appl
39	48.5	21.6	490	9	US-10-029-180-76	Sequence 76, Appl
40	48.5	21.6	635	9	US-09-877-650-15	Sequence 15, Appl
41	48.5	21.6	635	9	US-10-166-332A-2	Sequence 2, Appl
42	48.5	21.6	625	10	US-09-871-856-15	Sequence 15, Appl
43	48.5	21.6	1647	9	US-09-824-574-4	Sequence 4, Appl
44	48.5	21.6	1724	9	US-09-964-899-43	Sequence 43, Appl
45	48	21.3	75	10	US-09-864-761-37251	Sequence 37251, A

#### ALIGNMENTS

##### RESULT 1

US-09-738-626-5062  
; Sequence 5062, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, MOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patent ver. 3.0  
; SEQ ID NO 5062  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5062

Query Match 27.6% Score 62; DB 9; Length 291;

Best Local Similarity 47.8%; Pred. No. 4.8;

Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 6 LDCPTGPGVAIFGPGVMPQ 28

DB 163 LSTPTGTAFAFGAGVLMPE 185

```

CURRENT APPLICATION NUMBER: US/09/736,969A
CURRENT FILING DATE: 2000-12-13
PRIORITY APPLICATION NUMBER: US 60/160,860
PRIORITY FILING DATE: 1999-10-21
PRIORITY APPLICATION NUMBER: US 60/162,498
PRIORITY FILING DATE: 1999-10-29
PRIORITY APPLICATION NUMBER: US 60/170,453
PRIORITY FILING DATE: 1999-12-13
PRIORITY APPLICATION NUMBER: US 60/176,195
PRIORITY FILING DATE: 2000-01-14
PRIORITY APPLICATION NUMBER: US 60/182,296
PRIORITY FILING DATE: 2000-02-14
PRIORITY APPLICATION NUMBER: US 09/547,276
PRIORITY FILING DATE: 2000-04-11
PRIORITY APPLICATION NUMBER: US 60/196,267
PRIORITY FILING DATE: 2000-04-11
PRIORITY APPLICATION NUMBER: US 60/196,460
PRIORITY FILING DATE: 2000-04-11
PRIORITY APPLICATION NUMBER: US 60/196,527
PRIORITY FILING DATE: 2000-04-11
PRIORITY APPLICATION NUMBER: US 60/196,528
PRIORITY FILING DATE: 2000-04-11
PRIORITY APPLICATION NUMBER: US 09/687,837
PRIORITY FILING DATE: 2000-10-13
PRIORITY APPLICATION NUMBER: US 60/240,503
PRIORITY FILING DATE: 2000-10-13
PRIORITY APPLICATION NUMBER: US 60/240,508
PRIORITY FILING DATE: 2000-10-13
PRIORITY APPLICATION NUMBER: US 60/240,539
PRIORITY FILING DATE: 2000-10-13
PRIORITY APPLICATION NUMBER: US 60/240,543
PRIORITY FILING DATE: 2000-10-13
NUMBER OF SEQ ID NOS: 153
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
LENGTH: 1220
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human CLASP-3
US-09-736-969A-14

Query Match          24.0%: Score 54; DB 10; Length 1220;
Best Local Similarity 47.8%: Pred. NO. 2.2e+02;
Matches 11: Conservative 3; Mismatches 9; Indels 0; Gaps 0.

QY      11 TGGPGVATFGGPGVPWPQPTQSM 33
        |||| | : | : | : | : | : |
DB      62 TGGPRAPWGSNPSRSTQAM 84

RESULT 4
US-09-736-960-14
Sequence 14, Application US/09736960
Patent No. US20020102267A1
GENERAL INFORMATION:
APPLICANT: Lu, Peter
APPLICANT: Garman, Jonathan David
APPLICANT: Candia III, Albert Frederick
APPLICANT: Arbor Vita Corporation
TITLE OF INVENTION: CLASP-5 Transmembrane Protein
FILE REFERENCE: 020054-000511US
CURRENT APPLICATION NUMBER: US/09/736,960
CURRENT FILING DATE: 2001-09-20
PRIORITY APPLICATION NUMBER: US 60/160,860
PRIORITY FILING DATE: 1999-10-21
PRIORITY APPLICATION NUMBER: US 60/162,498
PRIORITY FILING DATE: 1999-10-29
PRIORITY APPLICATION NUMBER: US 60/170,453
PRIORITY FILING DATE: 1999-12-13
PRIORITY APPLICATION NUMBER: US 60/176,195
PRIORITY FILING DATE: 2000-01-14
PRIORITY APPLICATION NUMBER: US 60/182,296

```



```

; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 1220
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human CLASP-3
US-09-736-960-14
```

```

Query Match          24.0%; Score 54; DB 10; Length 1220;
Best Local Similarity 47.8%; Pred. No. 2.2e+02;
Matches 11; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
```

```
QY 11 TGGPGVAIFGCGPVPWPOFTQSM 33
      ||| | | | | | | | | | | |
Db 62 TGGPKAAPMGNSPSPSAESTQAM 84
```

```

RESULT 5
US-09-736-968A-107
; Sequence 107, Application US/09736968A
; Patent No. US20020169283A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-7 Transmembrane Protein
; FILE REFERENCE: 020054-000611US
; CURRENT APPLICATION NUMBER: US/09/736,968A
; CURRENT FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
```

```

; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 107
; LENGTH: 2090
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human CLASP-3
US-09-736-968A-107
```

```

Query Match          24.0%; Score 54; DB 9; Length 2090;
Best Local Similarity 47.8%; Pred. No. 3.8e+02;
Matches 11; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
```

```
QY 11 TGGPGVAIFGCGPVPWPOFTQSM 33
      ||| | | | | | | | | | | |
Db 933 TGGPKAAPMGNSPSPSAESTQAM 955
```

```

RESULT 6
US-09-736-969A-93
; Sequence 93, Application US/09736969A
; Patent No. US20020068302A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-4 Transmembrane Protein
; FILE REFERENCE: 020054-000411US
; CURRENT APPLICATION NUMBER: US/09/736,969A
; CURRENT FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 153
```



PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 41973  
LENGTH: 91  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL135747.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.3  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9  
OTHER INFORMATION: SWISSPROT HIT: P24054, EVALUATE 1.20e+00  
OTHER INFORMATION: EST\_HUMAN HIT: BF330102.1, EVALUATE 2.00e-22  
US-09-864-761-41973

Query Match 23.8%; Score 53.5; DB 10; Length 91;  
Best Local Similarity 52.6%; Pred. No. 15;

Matches 10; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 20 GCGVPMPOFTQSMY-YTL 37  
||| |||:|:| |:  
Db 16 GCGVCPMPRFSEPPDSHTL 34

RESULT 10  
US-09-823-240-9  
Sequence 9, Application US/09823240  
Patent No. US20020048813A1  
GENERAL INFORMATION:  
APPLICANT: Frank B. Gertler  
APPLICANT: James E. Bear  
APPLICANT: Jurgen Wehlend  
APPLICANT: Joseph Loureiro  
TITLE OF INVENTION: Methods and Products for Regulating Cell  
FILE REFERENCE: M0656/7064 (HCL)  
CURRENT APPLICATION NUMBER: US/09/823,240  
PRIOR FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: 60/194,564  
PRIOR FILING DATE: 2000-04-03  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 9  
LENGTH: 684  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-09-823-240-9  
Query Match 23.8%; Score 53.5; DB 10; Length 684;  
Best Local Similarity 52.2%; Pred. No. 1.3e+02;

Matches 12; Conservative 1; Mismatches 7; Indels 3; Gaps 1;  
QY 13 GCGVAFRGCGVPMPOFTQSMY 35  
||| ||| ||| ||| |||  
Db 351 GPG---YGGPVPPPPQQAENPY 370

RESULT 11  
US-10-023-282-740  
Sequence 740, Application US/10023282  
Publication No. US2003092893A1  
GENERAL INFORMATION:  
APPLICANT: Young et al.  
TITLE OF INVENTION: 207 Human Secreted Proteins  
FILE REFERENCE: P2007P1  
CURRENT APPLICATION NUMBER: US/10/023,282  
PRIOR FILING DATE: 2001-12-20  
EARLIER APPLICATION NUMBER: 09/205,258  
EARLIER FILING DATE: 1998-12-04  
EARLIER APPLICATION NUMBER: PCT/US98/11422  
EARLIER FILING DATE: 1998-06-04  
EARLIER APPLICATION NUMBER: 60/048,885  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,375  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,881  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,880  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,896  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,020  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,876  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,895  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,884  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,894  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,971  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,882  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,899  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,893  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,900  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,901  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,892  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,915  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,019  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,970  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,972  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,916  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,373  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,875  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,374  
EARLIER FILING DATE: 1997-06-06

```

? EARLIER APPLICATION NUMBER: 60/048,917
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,949
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,974
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,883
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,897
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,898
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,962
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,963
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,877
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,878
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/070,923
? EARLIER FILING DATE: 1997-12-18
? EARLIER APPLICATION NUMBER: 60/092,921
? EARLIER FILING DATE: 1998-07-15
? EARLIER APPLICATION NUMBER: 60/094,657
? EARLIER FILING DATE: 1998-07-30
? NUMBER OF SEQ ID NOS: 1227
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 740
? LENGTH: 81
? TYPE: PRT
? ORGANISM: Homo sapiens
US-10-023-282-740

```

```

Query Match      23.6%; Score 53; DB 9; Length 81;
Best Local Similarity 44.4%; Pred. No. 16;
Matches 12; Conservative 4; Mismatches 9; Indels 2; Gaps 1;

QY      12 GGGVAIFGCGPVPQPT--QSMRYT 36
      111: 11111 11111
Db       8 GPGGNNMGPGGRWPMPNTNNSIPYS 34

```

```

RESULT 12
US-10-023-282-739
? Sequence 739, Application US/10023282
? Publication No. US20030092893A1
? GENERAL INFORMATION:
? APPLICANT: Young et al.
? TITLE OF INVENTION: 207 Human Secreted Proteins
? FILE REFERENCE: P2007P1
? CURRENT APPLICATION NUMBER: US/10/023,282
? CURENT FILING DATE: 2001-12-20
? EARLIER APPLICATION NUMBER: 09/205,258
? EARLIER FILING DATE: 1998-12-04
? EARLIER APPLICATION NUMBER: PCT/US98/11422
? EARLIER FILING DATE: 1998-06-04
? EARLIER APPLICATION NUMBER: 60/048,885
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/049,375
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,881
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,880
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,896
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/049,020
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,876
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,895
? EARLIER FILING DATE: 1997-06-06

```

```

? EARLIER APPLICATION NUMBER: 60/048,884
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,894
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,971
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,964
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,882
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,899
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,893
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,900
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,901
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,892
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,915
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/049,019
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,970
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,972
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,916
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/049,373
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,875
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/049,374
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,917
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,949
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,974
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,883
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,897
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,898
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,962
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/070,923
? EARLIER FILING DATE: 1997-12-18
? EARLIER APPLICATION NUMBER: 60/092,921
? EARLIER FILING DATE: 1998-07-15
? EARLIER APPLICATION NUMBER: 60/094,657
? EARLIER FILING DATE: 1998-07-30
? NUMBER OF SEQ ID NOS: 1227
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 739
? LENGTH: 114
? TYPE: PRT
? ORGANISM: Homo sapiens
US-10-023-282-739

```

```

Query Match      23.6%; Score 53; DB 9; Length 114;
Best Local Similarity 44.4%; Pred. No. 22;
Matches 12; Conservative 4; Mismatches 9; Indels 2; Gaps 1;

```



Matches 12; Conservative 4; Mismatches 9; Indels 2; Gaps 1;  
QY 12 GCGVAIFGCGPVPPOFT--QSMPT 36  
| | | : | | | | : | | :  
Db 81 GMPGMNMGCGGRWPNTNANSTIPYS 107

RESULT :5  
US-10-023-282-738  
; Sequence 738, Application US/10023282  
; Publication No. US20030092893A1  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 207 Human Secreted Proteins  
; FILE REFERENCE: P2007P1  
; CURRENT APPLICATION NUMBER: US/10/023,282  
; EARLIER APPLICATION NUMBER: 2001-12-20  
; EARLIER APPLICATION NUMBER: 09/205,258  
; EARLIER FILING DATE: 1998-12-04  
; EARLIER APPLICATION NUMBER: PCT/US98/11422  
; EARLIER FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,885  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,375  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,881  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,880  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,896  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,020  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,876  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,895  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,884  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,894  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,971  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,964  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,882  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,899  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,893  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,900  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,901  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,892  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,915  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,019  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,970  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,972  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,916  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,373  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,875  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,374  
; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,917  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,949  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,974  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,883  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,897  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,898  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,962  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,963  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,877  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,878  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/070,923  
; EARLIER FILING DATE: 1997-12-18  
; EARLIER APPLICATION NUMBER: 60/092,921  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/094,657  
; EARLIER FILING DATE: 1998-07-30  
; NUMBER OF SEQ ID NOS: 1227  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 738  
; LENGTH: 235  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-282-738

Query Match 23.6%; Score 53; DB 9; Length 235;  
Best Local Similarity 44.4%; Pred. No. 49;  
Matches 12; Conservative 4; Mismatches 9; Indels 2; Gaps 1;  
QY 12 GCGVAIFGCGPVPPOFT--QSMPT 36  
| | | : | | | | : | | :  
Db 81 GMPGMNMGCGGRWPNTNANSTIPYS 107

Search completed: May 20, 2003, 18:00:04  
Job time : 21.2747 secs